

Article

Quantitative Analysis and Performance Evaluation of Target-Oriented Replication Strategies in Cloud Computing

Quadri Waseem ¹, Wan Isni Sofiah Wan Din ², Sultan S. Alshamrani ^{3,*} , Abdullah Alharbi ³ and Amril Nazir ⁴ 

¹ AnalytiCray, No 2-16, Jalan Pandan Prima 2, Dataran Pandan Prima, Kuala Lumpur 55100, Malaysia; qwaseem@analyticray.com

² Faculty of Computing, Universiti Malaysia Pahang, Gambang, Pahang 26300, Malaysia; sofiah@ump.edu.my

³ Department of Information Technology, College of Computer and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; amharbi@tu.edu.sa

⁴ College of Technological Innovation, Abu Dhabi Campus, Zayed University, P.O. Box 144534, Abu Dhabi, United Arab Emirates; mohd.nazir@zu.ac.ae

* Correspondence: susamash@tu.edu.sa

Abstract: Data replications effectively replicate the same data to various multiple locations to accomplish the objective of zero loss of information in case of failures without any downtime. Dynamic data replication strategies (providing run time location of replicas) in clouds should optimize the key performance indicator parameters, like response time, reliability, availability, scalability, cost, availability, performance, etc. To fulfill these objectives, various state-of-the-art dynamic data replication strategies has been proposed, based on several criteria and reported in the literature along with advantages and disadvantages. This paper provides a quantitative analysis and performance evaluation of target-oriented replication strategies based on target objectives. In this paper, we will try to find out which target objective is most addressed, which are average addressed, and which are least addressed in target-oriented replication strategies. The paper also includes a detailed discussion about the challenges, issues, and future research directions. This comprehensive analysis and performance evaluation based-work will open a new door for researchers in the field of cloud computing and will be helpful for further development of cloud-based dynamic data replication strategies to develop a technique that will address all attributes (Target Objectives) effectively in one replication strategy.

Keywords: cloud computing; dynamic replication strategies; target objective-based replication strategies



Citation: Waseem, Q.; Wan Din, W.I.S.; Alshamrani, S.S.; Alharbi, A.; Nazir, A. Quantitative Analysis and Performance Evaluation of Target-Oriented Replication Strategies in Cloud Computing. *Electronics* **2021**, *10*, 672. <https://doi.org/10.3390/electronics10060672>

Academic Editor: Javid Taheri

Received: 3 February 2021

Accepted: 9 March 2021

Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last few years, cloud computing has shown a significant impact in the field of storage systems. It is recognized as web-based administration of configurable, parallel, and adaptive systems and has advanced as a most recent approach for accessing, managing, and controlling the massive, distributed data at various geographical areas. The main purpose of cloud computing is to provide a simplified and proficient on-demand network access, along with service to a pool of shared virtualized processing assets based on a pay-as-you-go agreement [1–4]. Besides providing data availability, it additionally improves load balancing, fault tolerance, and scalability. Moreover, it minimizes the job execution time, bandwidth consumption, and performance. The services offered by the cloud incorporate infrastructure flexibility, cost control, faster application deployment, data adaptation of cloud resources to real needs, and improve profitability. In distributed data centers, there is a huge demand to store plentiful data on cloud foundations due to the integration of computer networks, servers, storage, and numerous related programming schemes [5]. The expanding quantities of cloud-facilitated applications that are fueled by the cloud-facilitated database systems are generating and consuming a tremendous volume of data at

an extraordinary scale. Hence, the reason to use the distributed computing worldview is to simplify and overcome the time-consuming processes of software deployment, equipment provisioning, and purchasing [6]. Furthermore, these cloud data centers successfully furnish high-performance computing, along with accessibility, scalability, availability, adaptability, quick deployment, cost adequacy, real-time variations, and efficiency for the premier data storage [7,8].

Apart from cloud computing, which is well-known for offering scalable computing and storage services [9], big data technology is also gaining the momentum around the world to help companies gain a better understanding of their data and make better decisions. Big data helps users to process distributed queries through several databases using commodity computing, while Hadoop, a class of distributed data-processing systems, provides the underlying engine of cloud computing. The cloud environment is the best option for addressing storage problems due to the huge volume of data and complex data computation in big data technologies. Big data makes use of cloud-based distributed storage technologies [10,11]. In the computing world, more and more data-intensive applications are being created [12]. In big data, Hadoop's file systems are handled by the Hadoop Distributed File System (HDFS), which also manages the storage resources [13].

Big data and cloud computing are both interrelated technologies. Besides providing various opportunities, facilities, and services conjointly technologies, these technologies face various technical issues and challenges. Hence, an attention is needed for their smooth working and for their performance enhancement. Various research works have been done related to big data and cloud computing issues, which address the storage part, integration part, their complexities, and their future prospective directions [14–20]. The Paper [21] explores the issues, the challenges, and the future directions of big data in cloud computing. Some of the important key research issues and challenges include availability, data transformation, data quality, scalability, data heterogeneity, legal issues, privacy, data integrity, and regulatory governance, which are all discussed in detail [21].

Due to heterogeneity of large data systems. The key problem with big data and its storage is data availability. Users must have access to data at all times, regardless of the time. The most efficient way to meet this desire is to provide the most reliable replication methods that will ensure business continuity effectively [10]. As a result, replication is the best approach to deal with all such issues in cloud computing, while keeping in mind the heterogeneous existence of big data and the storage problem of big data. Some of the application of replication and its related domains have been explored and elaborated in References [22,23].

1.1. Background

1.1.1. Data Replication in Clouds

Although cloud storage and computing have been extensively acknowledged by numerous organizations, still there are many concerns when we talk about the failures created due to disasters. For the best storage of data on the cloud, the system should have the best replication strategies to accomplish the objective of saving the information in the case of any disaster events [8]. Therefore, the performance and availability can be increased by replicating the data to various locations in the cloud system where exactly the applications are controlled and executed [24]. Hence, the basic prerequisites of cloud storage systems rely on the replication strategy and its related data consistency methods [25]. There is always a demand for data replication in distributed large storage systems.

1.1.2. The Need for Data Replication

Most organizations are utilizing distributed computing to store and access data remotely. The stored data needs to be reserved to offsite to ensure simple and easy recovery during events of downtime. Data replication permits organizations to scale their offsite storage quickly for faster backup and recovery. Data replication is a process of storing multiple copies of similar data on various storage devices or storage nodes. If, at any node,

there is an update/write operation, a similar update should be immediately passed to other replicas, too. The main idea is to recover the lost data by utilizing these replicated copies from the cloud [26]. Data replication is considered a performance-enhancing technique for cloud storage frameworks that generally has been used and adopted by large-scale cloud storage systems. In large-scale cloud storage systems, it is the only solution that provides the data availability along with performance in case of occurrence of disaster(failures). By utilizing these numerous replicated copies, data replication guarantees high information sharing and access latency, along with improved system load balancing. One of the biggest advantages of data replication is its consistency in decreasing the response time and improving the reliability [3,27]. Other advantages include accelerating the data access, reducing access latency, least network delays (user waiting time), and bandwidth usage (cloud system bandwidth capacity utilization) [5,28]. Consequently, data replication is used in the clouds for upgrading the performance (e.g., read and write delay) of applications that access the data [8]. However, the risk of node failure in cloud storage frameworks within a data-intensive application is around the clock [25]. Proper implementation and execution of data-replication mechanisms over the cloud services will promote the availability, fault-tolerance, and failure recovery [29]. Therefore, keeping the data at more than one site will increase the availability, and the request can discover the data close to the site where the request originally began, subsequently limiting the service request time and improving the performance of the system in general.

1.1.3. Research Motivation

There has been extensive research going on to optimize various types of dynamic replication strategies. We try to analyze and evaluate the target-oriented replication strategies in large-scale cloud storage systems based on target objectives which are represented through various attributes correspondingly discussed in different previous works [30–32]. The attributes associated with these target objectives are Availability, Reliability, Performance (Storage Space, Storage cost, Bandwidth Consumption, Response time), Fault Tolerance, Load Balance Scalability, Elasticity, Consistency, and Cost. The main motivation for this research is to discover the target objectives related to dynamic replication strategies and elaborate on their utilization in each replication strategy.

1.1.4. Paper Organization

This paper is organized as follows. In Section 2, we present the review methodology. In Section 3, we presented a taxonomy for data replication strategies. In Section 4, we present a taxonomy for dynamic cloud computing replication strategies and target-oriented taxonomy for dynamic replication in cloud computing along with relationships of target objectives based on attributes. The section also contains the quantitative analysis summary. In Section 5, we present a comparison and evaluation in details. In Section 6, we present challenges of replication strategies in clouds. In Section 7, we present the least addressed target objective of dynamic replication strategies in clouds, their challenges, issues, and future research directions. In Section 8, we conduct our discussion. In the final section, we present our concluding remarks.

2. Research Methodology Used

In our first phase of research methodology, we select the research papers for our critical analysis and evaluation through searching various types of databases. In the second phase, we include and exclude the papers based on their title, abstract, and main content. The next phase followed was by checking the accepted papers against each formulated research objectives. Finally, in the last phase of our research methodology, which was based on reading the full content, the main papers were collected for our quantitative analysis and performance evaluation.

In this section, we discuss the research questions related to our research, source of information, service criteria, quality assessment, and review phases.

2.1. Research Questions

In this section, we present the research questions we adopted in our critical analysis and performance evaluation. The motivation behind each research question is mentioned accordingly shown in Table 1.

Table 1. Research questions.

No.	Research Questions (RQ)	Motivation behind Each Research Question
1	RQ1 What are the main areas of research related to replication techniques especially in cloud computing?	Mainly, the aim is to identify and evaluate the various data replication strategies related to articles/studies based on various replication strategies of previously published and mention their importance in cloud computing environments.
2	RQ2 What are the main target objectives, a replication strategy should possess?	Here, we discuss various dynamic replication strategies based on different categories especially the target-based dynamic replication strategies to understand the need for each replication strategy.
3	RQ3 What are the most used research replication strategies and how are they applied in the cloud replication area?	
4	RQ4 What attributes a replication strategy might consider meeting the target objectives?	This research will aim to provide different target-based objectives of dynamic replication strategies and their dependent attributes for best optimization. The relationship will help to understand the utilization of different algorithms for the best performance.
5	RQ5 What is the relationship between target objectives with their concern parameters?	
6	RQ6 What is the relationship between different parameters and what are their metrics?	Various research papers need to be identified from different replication strategy categories to reveal vital research problems. The research will display the quantitative analysis for performance evaluations of target-oriented replication strategies in cloud computing. There is a need to develop a technique that will address all attributes (Target Objectives) effectively in one replication strategy.
7	RQ7 What are the main metrics used for performance evaluation purposes?	
8	RQ8 What are the key results obtained?	This research also aims to identify the main issues and challenges of existing target-oriented replication strategies along with future directions to ensure optimal services. Various questions discussed here will help in the identification of future research areas.
9	RQ9 What are the main challenges and open issues of replication in cloud computing?	

2.2. Sources of Information

We searched for various digital library sources (Scopus, Web of Science, Google Scholar, etc.) to find the relevant papers related to cloud-based replication. We searched for journals and conferences and for books to extract the relevant research papers. The following databases has been used in our search: Springer, ScienceDirect, Scopus, Google Scholar, ACM Digital Library, IEEE Xplore, and Taylor & Francis.

2.3. Search Criteria

We formulated the keywords to search in the above-mentioned databases using the specific keywords “Data Replication Strategies/Techniques” and “Cloud Computing”. In this research, we use the title and abstract of the research papers to get our results. We tried searching various related keywords that matched our target results, like “cloud-based replication strategies” and dynamic data replication strategies. Then, our process of searching the articles was based on adding the “Target Objective” prefix, like “Target objective cloud-based replication strategies” or its synonym “target-oriented cloud-based replication strategies”. We also searched by using various parameters/attributes of cloud computing replication, for example, “Performance” in a way like Performance Analysis of Data Replication Strategies in Cloud”.

2.4. Quality Assessment

On searching for articles related to our topic, we applied the inclusion and exclusion criteria mentioned below:

For Inclusion, we followed:

1. Clearly describes target objectives of replication strategies for cloud computing.
2. Peer-reviewed articles in the English language.
3. Articles published in reputable journals, conferences, and magazines.
4. Articles published from 2011 to 2019.

For Exclusion, we followed:

1. Does not focus on dynamic replication strategies in the cloud.
2. Articles that are not related to the research questions.
3. Articles whose full text is not available.
4. Articles that have common challenges and references.

2.5. Review Phases

After defining the search keyword, our four-stage review phases are summarized as:

- First of all, the articles were searched based on defined keywords (mentioned in search criteria) and were initially found to be 109 articles in total.
- Then, articles were excluded that do not meet inclusion, exclusion criteria. This criterion minimizes our article search to 53 articles.
- Then, research question objectives were used for further filtration of articles. This criterion also minimizes our article search to 28.
- Finally, articles were evaluated based on full paper reading and the total papers finalized for this research were 22.

For RQ1–RQ5 and RQ8 and RQ9, we collected a total of 108 papers, which include the related surveys also. For RQ 6 and RQ 7, we collected a total of 22 papers.

In this paper, we initially introduce a taxonomy of replication strategies, along with their related survey of surveys. Then, we provide the taxonomy of dynamic cloud replication strategies, along with their related survey of surveys. At last, our focus of this paper is target-oriented replication strategies and their taxonomy, along with a detailed investigation.

3. Data Replication Strategies

In the last few years, there has been a huge contribution from many researchers, scientists, and academicians in the field of data replication. The contribution not only gives the optimal solution to the basic issues of the replication strategies but also provides a smooth way to implement these replication strategies in different distributed architectures. The main intent is to get the benefit of replication strategies in various types of distributed architectures mentioned in Refs. [33,34], which include Distributed Database Management Systems, Peer to Peer Systems, Data Grid, Worldwide Web, Distributed Geographic Information Systems, and others, especially in Cloud Computing [35]. At present, many efforts have been utilized which have strengthened the replication strategies roots deep into the cloud computing architectures. On this subject, various researchers have contributed vigorously. Several researchers had contributed for the implementation-related issues, several contributed for optimization, and several researchers have provided reviews which include classification and taxonomy of replication strategies for cloud-based structures using different criteria.

Figure 1 depicts the taxonomy of data replication strategies.

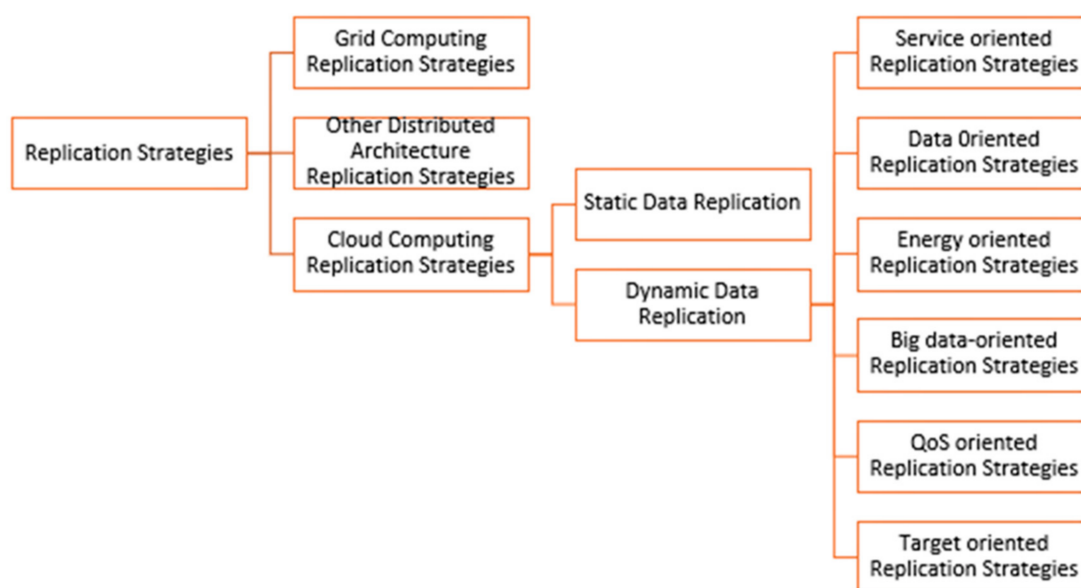


Figure 1. Taxonomy of data replication strategies.

In this section, we present a taxonomy of replication strategies based on a distributed architecture (shown in Figure 2), we have categorized the data replication into (1) grid computing replication strategies, (2) other distributed architecture replication strategies, and (3) cloud computing replication strategies.

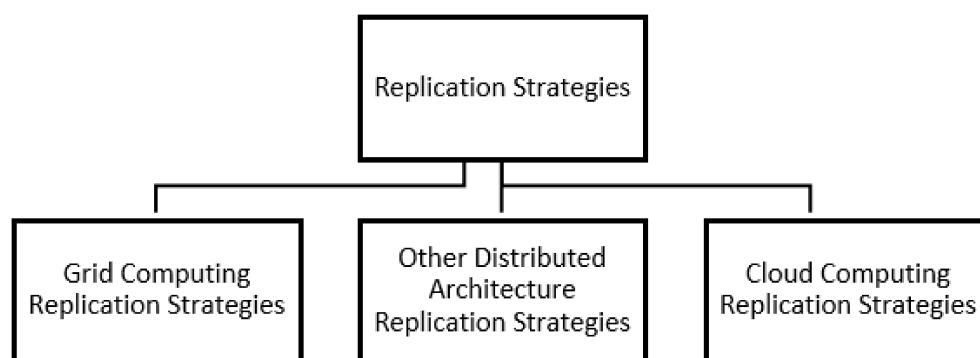


Figure 2. Demonstrates the classification of replication strategies based on a distributed architecture.

3.1. Grid Computing Replication Strategies

A data grid is a cluster of services that furnish smooth access, modification and transfer of a substantial amount of data over geographically distributed structures. Hence, massive storage resources are the basic requirements for the storage of data files. To furnish the storage of data files in these large storage systems, data replication makes a great impact by scaling back data time intervals and using fast network and storage resources efficiently for the efficient recovery [36]. Grid computing-based replications are utilized in various scenarios, and its research has a tremendous future scope.

Related Surveys

Various research efforts have been made for the replication strategies related to grid computing. We collected some of the reviews of various replication strategies for grid-based environments, for the sake of understanding the replication strategies concerning grid computing terminology.

Amjad Sher et al. [30] proposed an extensive review on grid computing replication strategies. They split the replication strategies into many categories based on nature and

architecture of data grid structures. The proposed survey was targeted to enhance and improve the data availability using dynamic replication strategies in data grids [30].

Hamrouni et al. [32] reviewed the data replication strategies and specifically stressed more on replica selection strategies. Replication selection strategies act as an important data management technique mostly used in data grid structures for the enhancement in network performance, file access patterns, user or job access behavior, and file correlations, as well as in prediction of future behavior. The strategies were discussed, along with advantages and disadvantages [32].

Naseera et al. [37] proposed a comprehensive survey on the issues and challenges involved in grid environment-based data replications with a focus on concerns, such as replica consistency, replica synchronization, and replica maintenance. This survey provides a general review on data replication related to various important aspects, namely replica creation/modification, replica selection, the optimal number of replicas, and replicas consistency. They also mentioned the limitations, as well as future enhancements [37].

Vashisht et al. [38] classified and analyzed various asynchronous replica consistencies which were classified based on different criterion, such as the level of abstraction, load balancing, update propagation, fault tolerance, topology, location, check-pointing, and many more strategies [38].

Tos et al. [39] presented a survey of the latest dynamic grid-based data replication strategies. The classification criteria for their strategy are based on target data grid architecture. Their work includes the survey of the strategies and their feature comparison using important metrics for evaluation [39].

Hamrouni et al. [40] presented a similar work for replication strategies in a grid computing domain particularly using data mining techniques. This study narrates the use of data mining techniques in grid-based setups to understand and evaluate historical data [40].

Mansouri et al. [41] proposed a survey which investigated to determine which attributes are assumed in each replication algorithm and which are declined. They represented the important factors to facilitate the future comparison of data replication algorithms and presented some interesting discussions about future works along with open research challenges [41].

Souravlas et al. [42] provided a general summary of latest strategies for replication based on selection criteria (geography, space or time) for data files to be replicated. Moreover, they mentioned the pros and cons of each strategy and evaluated the performance based on a bunch of parameters [42].

Some of the latest research to enhance the replication strategies in the field of grid computing are addressed in Ref. [43] for replica creation, Refs. [44,45] for replica placement, and Ref. [46] for distributed database systems.

3.2. Other Distributed Architecture-Based Replication Strategies

Replication strategies related to other distributed architecture include distributed database management systems, peer-to-peer systems, worldwide web, utility-based distributed systems, and many applications, like mobile systems, artificial intelligence, business application, etc. These systems are mostly concerned with the need basis and are application oriented. There is always a strong connection between the applications and the distributed systems. Several applications are developed based on system needs which keep on growing with time. A lot of research has been done beginning with a simple creation of architecture based on the number of requests, initiated by a client to a server. Such vital architectures are unable to handle large numbers of requests, and there is always a performance constraint to maintain the response time and to efficiently use the network bandwidth. To some extent, the mobile agents strive to control these above discussed demerits but could not succeed fully to cover up the growing demand and technology setups [47]. Other architecture-base replications strategies are utilized in various distributed firms and hence more efforts should be used for their performance.

Although peer-to-peer (P2P systems) are mainly designed for read-only database applications, while as others deal with transactional queries related to databases, data grid systems deal with read-only queries. The benefits of the replication in read-only database applications can be neutralized by the overhead of maintaining continuity among multiple replicas if the application needs to process updated queries [48]. The latest roles of various other architecture-based applications, their utilization in various domains, and analytics can be found in Reference [49]. Replication related to applications like mobile systems, artificial intelligence, business application, etc., is mostly dependent on storage utilization. The application domains decide which storage systems to use and what the processing techniques should be, while keeping the storage restriction in view.

Related Surveys

Various research efforts have been made for the replication strategies related to other distributed architecture computing. We collected some of the reviews of various replication strategies for other distributed architecture-based environments, for the sake of understanding the replication strategies with respect to other distributed architecture terminology, like (peer-to-peer systems) p2p, (database management systems) DBMS, mobile computing, etc.

Sushant Goel et al. [34] presented an extensive review of distributed storage and data distribution systems, where they split the distributed systems based on their architecture into four subclasses, namely (a) Distributed database management systems, (b) Peer-to-peer Systems, (c) Data grids, and (d) Worldwide web. Furthermore, their contribution also includes the further classification of the above four subclasses in detail [34].

Whereas the review of Spaho et al. [50] presented a survey of p2p systems, the proposed survey is based on the classification of replica placement strategies by utilizing the criteria of site selection and replica placement. These two criteria provide the depth in comprehensive classification of P2P systems [50].

Some of the latest research to enhance the replication strategies in other distributed architecture are addressed in Reference [51] for Cloud-P2P environments, Reference [52] for document-oriented (Not only SQL) NoSQL (Not only SQL) systems, Reference [53] for replica selection in Internet of Things (IoT), Reference [54] for cost aware heterogeneous cloud data centers, and Reference [55] for mobile ad hoc networks structures.

3.3. Cloud Computing Replication Strategies

In a cloud-based replication, the data files are split into multiple blocks over the distributed network. The aim is to have multiple copies (replicas) of the same data at various distributed data nodes. However, the network dependency factor within the data-intensive application causes the node failures in the large cloud storage system. These network factors include (e.g., bandwidth, node failure and untrustworthy networks). If a node holding a data file fails to work, then the whole data file will be gone. Therefore, there is always a need for data availability [25].

Machine learning refers to a collection of algorithms that can detect patterns in data and predict outcomes in the event of a decision. Machine learning algorithms have been used to avoid or detect attacks and security problems, including cloud vulnerabilities, in a variety of ways [56]. The use of machine learning and its applications in cloud computing and related environments has been discussed in some of the most recent related works [57–61].

Different users exchange sensitive data over the cloud in cloud computing, and failures are possible. As a result, data fragmentation and replication algorithms can help improve data protection. As a result, the idea of safe data replication (SDR) was developed, in which attackers are unable to determine the positions of replicas and the replication process is secure [62]. Machine Learning techniques are used in replication to secure the clouds [63]. In Reference [64], the authors use machine learning to implement a multi-objective optimization data placement strategy in large-scale networked storage systems

that considers data protection and retrieval time. As a result, it ensures that the replication process can run faster and be more secure.

Data replication techniques in clouds are broadly labeled into two basic categories, which include static replication mechanism and dynamic replication mechanism [5], and their summary is presented in Table 2.

Table 2. Static data replication versus dynamic data replication.

Static Data Replication		Dynamic Data Replication	
Brief Description	In static data replication, a predefined set of replicas and host nodes are the key factor to achieve the data distribution at multiple sites. It determines the replica node locations at design phase.	Brief Description	In dynamic data replication, the key factor to achieve the data distribution at multiple sites is its automatic/adaptive nature of creating and omitting the replicas, based on user behavior and network topology. It determines the locations of replicas nodes at a run time.
Key Features	The static replication strategy accompanies deterministic policies in which the host nodes, the replica numbers are pre-decided and very much characterized.	Key Features	The dynamic strategies by default built and removes the replicas based on storage capacity changes, bandwidth and user access patterns (adaptive in nature).
	The static replication strategies are always simple to implement because number of replicas is constant.		These strategies are not easy to implement because number of replicas is variable (based on heterogenous workload).
	There is a need to support the random policy to keep the number of active service replicas at the maximum.		Being intelligent in nature, dynamic data replication is developed to make smart choices to choose the location of the data based on current available information.
Drawbacks	They are used less in real scenarios because of their predetermined nature.	Drawbacks	It is very difficult to control and accumulate the runtime information of all the data nodes in a complex cloud setup.
	The more active service replicas guarantee more performance, but performance cannot be obtained at a high operation cost.		It takes a lot of effort to maintain the data file consistency effectively.

Related Surveys

Various research efforts have been made for the replication strategies related to cloud computing. We collected some of the reviews of various replication strategies for cloud-based environments for the sake of understanding the replication strategies concerning cloud computing terminology.

We have summarized some of the replication strategies in cloud computing, along with some basic categories based on various types of classifiers.

Milani et al. [35] presented a detailed investigation of data replication strategies in cloud computing environment. The authors examined the data replication mechanisms in a cloud environment and studied the features and challenges, as well as addressed the relevant issues in data replication. Additionally, they provide a detailed comparison of the data replication strategies in cloud computing [35].

Tabet et al. [65] proposed a review of data replication in clouds systems. They divided the data replication of clouds into various categories based on different taxonomies as

objective function (static and dynamic), (replica factor optimal number and dynamic adjustment), (customer and provider centric), and (proactive and workload balancing) [65].

Bhuvaneswari et al. [66] proposed an extensive general review of data replication mechanisms for distributed systems. The review was broadly split into two main categories, consisting of dynamic and static replications irrespective of their architecture types, like grid, cloud, or network [66].

Some of the latest research to enhance the replication strategies in the field of cloud computing are addressed in Reference [67] for dynamic cost-aware replication, Reference [68] for cloud/edge based infra-structures, Reference [69] for mobile edge computing (MEC), Reference [70] for replication placement for in geographically distributed clouds, and Reference [71] for replication management in the cloud.

Due to predefined sets of replicas and host nodes which are determined at design phase, static replication strategies are used less in real scenarios. To overcome these hurdles, dynamic replication has emerged as a best alternative due to their adaptive nature of creating and omitting the replicas based on user behavior and network topology. These attractive characteristics had motivated us to select the dynamic replication strategies as our topic for further research.

4. Dynamic Cloud Computing Replication Strategies Taxonomy

In this section, we provide a taxonomy of dynamic cloud computing replication strategies and categorize them based on their service and tasks (shown in Figure 3). Dynamic cloud data replication strategies are divided into following subcategories:

1. Service-oriented replication strategies;
2. Data-oriented replication strategies;
3. Energy-oriented replication strategies;
4. Big data-oriented replication strategies;
5. Quality of service (QoS)-oriented replication strategies; and
6. Target-oriented replication strategies.

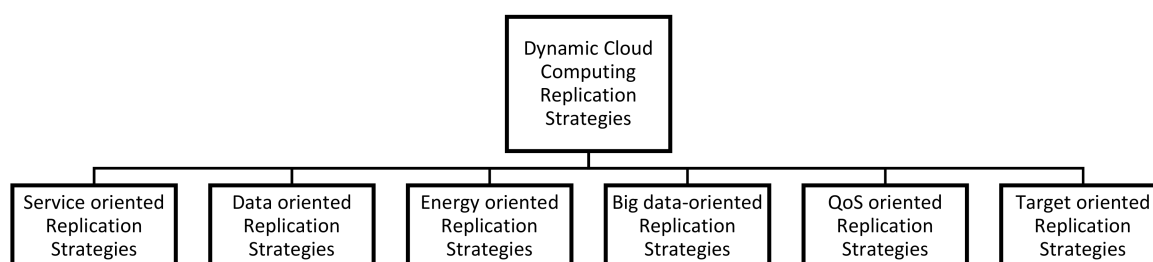


Figure 3. Demonstrates the classification of dynamic cloud computing replication strategies based on services and tasks.

4.1. Service-Oriented Replication Strategies

Service replication supports the non-functional requirement of services, in accordance with the understanding of Service-Level-Agreements (SLA). These services include data availability, response time, and data reliability [72]. In service-oriented replication for cloud systems, the service replicas utilize the storage resource, as well as other resources, such as central processing unit-CPU, memory, network, bandwidth, etc. The cost of replication and service dependencies are always high [73]. Therefore, service-oriented replication strategies are generally expensive in nature.

Related Surveys

Various research efforts have been made for the replication strategies related to service-oriented computing. We collected some of the reviews of various replication strategies

for service-oriented environments for the sake of replication strategies concerning service-oriented computing terminology.

Slimani et al. [73] presented an extensive review and classification of replication approaches as SoR (Service-oriented Replication) strategies and DoR (Data-oriented Replication) in cloud computing paradigm based on replicating the service or the underlying data. The proposed survey reviewed the latest replication techniques for the basic purpose to achieve high availability and QoS in cloud computing paradigms [73].

Mohamed et al. [74] presented a review of service-based replication, their challenges, their techniques, their types, and their algorithms in different distributed setups (service-oriented architecture (SOA), cloud, and mobile). Additionally, they also examined and explained the participation of replication in promoting various QoS attributes, such as availability, reliability, scalability, performance, and security [74].

Some of the latest research in the field of service-oriented replication for cloud computing are addressed in Reference [72] for replica provisioning policy, Reference [75] for dependency aware dynamic replication, Reference [76] for replicas placement, and Reference [77] for consistency-based replication.

4.2. Data-Oriented Replication Strategies

The process of replicating the underlying data is a commonly used technique to avoid failures and is commonly known as data-oriented replication. The cost of replicating a file is much lower than replicating a service. Hence, data-oriented replication strategies are cheaper as compared to service-oriented strategies. The data-oriented replication strategies have been subdivided into three major groups based on type of cloud application workload. The first one is data-intensive workload-based, the second one is computationally intensive workload-based, and the third one is balanced workload-based [73]. While comparing data-oriented replications with service-oriented replication, the data-oriented replications are easy to implement and are more performance-oriented.

Related Surveys

Various research efforts have been made for the replication strategies related to data-oriented computing. We collected some of the reviews of various replication strategies for data-oriented environments for the sake of understanding the replication strategies concerning data-oriented computing terminology.

Milani et al. [5] presented a work that specifically categorized the replication strategies in cloud systems into two main categories: (1) static replication strategies and (2) dynamic replication strategies. Static replication strategies choose the location of replication nodes and creation of replica during the design phase (predetermined), while dynamic replication strategies choose the replication nodes and creation of replicas at a run time (automatically) under the changes in the user access pattern, bandwidth, and storage capacity [5].

Malik et al. [7] presented a survey on data management and replication approaches. The focus of the survey is more on resource usage and QoS provisioning. They also analyzed the performance, advantages and disadvantages of data replication and data management in cloud-based setups. Furthermore, the paper discusses the issue and challenges related to consistency, load balancing, scalability, processing, and data placement [7].

Tabet et al. [65] presented a comprehensive survey of data replication for underlying data in cloud systems. The proposed survey is based on five dimensions. The first one is static versus dynamic, the second one is reactive versus proactive workload balancing, the third one is provider versus customer-centric, the fourth one is optimal number versus dynamic replica adjustment, and the last fifth one is the objective function-based [65].

Some of the latest research in the field of data-oriented replication for cloud computing are addressed in Reference [71] for replication management, as well as Reference [78] for replica placement.

4.3. Energy-Oriented Replication Strategies

Energy-oriented replication strategies are part of green computing. Green computing represents purifying the environment with a focus on storage, temperature, and energy. Recent research showed that large-scale data centers consumed a huge amount of electricity [79]. Therefore, for least energy consumption, the sum of active servers should be minimized, and the utilization level of replicas should be considered, although reducing the energy consumption and maintaining high computation capacity is done by implementing the replication strategies. However, the number of data replicas are directly proportional to energy consumption, which directly affects the performance and the cost of creating and maintaining new replicas [4]. Therefore, the primary issue is to decide the number of required replicas and their location.

Related Surveys

Various research efforts have been made for the replication strategies related to energy-oriented computing. We collected some of the reviews of various replication strategies for energy-oriented environments for the sake of understanding the replication strategies concerning energy-oriented computing terminology.

You et al. [80] provided a survey that gives a comprehensive understanding of the current level of energy efficiency related to surveys in cloud-related environments. Here, a survey on surveys of energy efficiency was performed based on five categories, which include the surveys on the energy efficiency of the whole cloud, of the certain levels in cloud, on a certain energy efficiency technique, on all energy-efficient strategies, and other energy efficiency-related surveys [80].

Ali et al. [81] presented a taxonomy of energy efficient techniques for cloud computing. The authors discuss the issues pertaining with huge energy consumption by cloud data centers. They presented a taxonomy of huge energy consumption issues, along with their solutions [81].

Some of the latest research in the field of energy-oriented replication for cloud computing are addressed in Refs. [4,78] for replication decision criteria, Reference [28] for communication delays, and Reference [82] for disk performance.

4.4. Big Data-Oriented Replication Strategies

The latest research shows that the cloud is the best solution for data-intensive applications. It is the only solution for optimal storage and provides terrific performance for huge data on distributed systems. Hence, a planned strategy between cloud and big data is needed to ensure consistent data accessibility without any disruption [83]. The recent research targets to provide data availability and maintain the performance of big data on clouds, even in case of disasters. Since the cloud distributes the huge, big data to various nodes either in the same data center or across many data centers on clouds [9,12,13,84].

Therefore, a reliable and efficient solution should be executed to overcome the failures using an optimal replication strategy in the cloud.

Related Surveys

Various research efforts have been made for the replication strategies related to big data-oriented computing. We collected some of the reviews of various replication strategies for big data-oriented environments for the sake of understanding the replication strategies concerning big data-oriented computing terminology.

Gopinath et al. [25] came up with a survey which includes the detailed survey related to replication and their implementation in big data domain, such as HDFS (Hadoop distributed file system). The survey gives the empirical evaluation and provides depth in the survey in the form of static and dynamic replication techniques [25].

Lalitha Singh et al. [83] introduced a related survey of cloud-based scientific workflows of various data placement strategies. Data placement strategies which use the big data are

studied in detail. The main purpose of the study is to improve the performance and the data movement cost [83].

Fazlina et al. [10] introduced a survey that emphasized more on the performance factors that classify the replication strategies into static and dynamic replication based on their metrics. The survey discussed the critical review along with the imperative details collected from various references. Moreover, they also discuss the gaps in replication strategies [10].

Mansouri et al. [85] presented a critical review with imperative details. They discussed the sudden move of data-intensive (big data) applications in connection with heterogeneous distributed computing systems for efficient data management. This work presents a complete review of data replication based on cloud computing and data grid computing [85].

Some of the latest research in the field of big data-oriented replication for cloud computing are addressed in Reference [86] for elastic replication, Reference [87] for predictive analysis-based replication, Reference [88] for dynamically replica adjustment, and Reference [89] for proactive data management.

4.5. QoS-Oriented Replication Strategies

QoS aware replication needs to allocate replicas by considering the Quality of Services (QoS) requirements of cloud, such as network delay, bandwidth, loss rate, etc. QoS provides the performance guarantee and other quality of vital services, such as availability, reliability, security, dependability, etc. Being directly associated with end-users and service providers, the QoS requirements are to deliver the services according to predefined agreements [90]. Many existing replication services [91] are designed for enhancing the system-oriented metrics rather than user-oriented metrics.

Related Surveys

Various research efforts have been made for the replication strategies related to QoS-oriented computing. We collected some of the reviews of various replication strategies for QoS-oriented environments for the sake of understanding the replication strategies concerning QoS-oriented computing terminology.

Saraswathi et al. [92] provided a detailed survey of data replication on the cloud environment, the proposed classification divides the data replications into QoS aware data replication and dynamic data replication strategies. The paper also mentioned the different applications have different quality-of-service (QoS) requirements and also concluded that it is very tough to maintain the common QoS during the running phase of the applications [92].

Zia et al. [93] provided a survey on various schemes about QoS issues. They analyzed the strengths and weaknesses based on their performance. This paper also investigated how the performance can be expanded by improving various segments like QoS and cost [93].

Some of the latest research in the field of QoS oriented replication for cloud computing are addressed in Reference [12] distributed cloud data placement [9], for cost-based replication, Ref. [94] for edge cloud-based replication, Ref. [95] for cost-based data replications and placement, and Refs. [96,97] for replica placement.

4.6. Target-Oriented Replication Strategies

Every single replication strategy consists of algorithms developed to meet contrasting objectives in certain environments. The main aim is to reform the divergent performance metrics. Depending on the area to be addressed, the algorithm would enhance various performance metrics, such as bandwidth usage, accuracy, response time, energy consumption, etc. [98,99]. Some of the real-time implementation of replication strategies stress more on fast response time and are used for big data domain, few others are implemented to reduce the data storage costs, and few are developed for transfer of workflow applications [10].

Each dynamic replication has its target, which represents its objectives. The basic target objectives of data replication strategies are known as primary objectives which include availability, reliability, and performance. The secondary objectives include fault tolerance and load balancing. Besides primary and secondary objectives, there are also tertiary objectives which are as much as important as that of primary and secondary objectives and must be addressed for efficiency and better performance. They include scalability, elasticity, consistency, and cost.

4.6.1. Taxonomy of Target Oriented Replication Strategies

In this study, we examine and collect the different types and categories of surveys related to cloud-based replication which incorporate the title, survey aim, perspective, target components, and year of publication. To the best of our knowledge, there is not a single detailed research about replication strategies in cloud environments which is systematic and comparative. Therefore, we conducted our research from the view of target objective-based philosophy and, hence, provide a systematic target-oriented taxonomy of dynamic replication strategies in the cloud. The available latest surveys include Refs. [65,73].

Before discussing the taxonomy of target-oriented replication strategies in detail, let us have a generalized look at the available literature (shown in Table 3).

Table 3. Literature review of target-oriented replication strategies.

Replication Strategy Basic Details	Replication Strategy	Advantage and Disadvantages
[100] Year 2011	Description: A cost-effective dynamic data replication strategy, namely (CIR), which is based on an incremental replication method with the aim to reduce the storage cost while maintaining the data reliability requirement. The approach calculating the replica creation to mention the storage duration.	Advantages: High data reliability, High availability, Low replication cost, and Low energy consumption. Disadvantages: High response time and Low load balancing.
[101] Year 2012	Description: A novel dynamic data replication strategy, namely (D2RS), which calculates a suitable number of copies based on evaluation and identification of popular data. Moreover, it also analyses and models various relationships accordingly.	Advantages: High availability, Low bandwidth consumption, and Low replication cost. Disadvantages: High user waiting time, Low speed data access, and Low load balancing.
[102] Year 2015	Description: A cost-effective data reliability mechanism, namely (PRCR), which is based on a generalized data reliability model. It works on a proactive replica checking approach to ensure the reliability of the data while maintaining the minimum number of replicas.	Advantages: Cost effective reliability, Less failure rates, Reduced storage space, and storage cost. Disadvantages: No reduction in response time and Low load balancing.
[26] Year 2012	Description: A replication, namely (Harmony), which handles the key issue in data management and provides the solution to deal with duplicate copies. The basic steps of the technique include determining the files for replication, time of replication, and deciding the final data location for replication.	Advantages: High availability and High performance. Disadvantages: High Downtown and High transactional cost.
[103] Year 2013	Description: A replication method, namely (Bismar), which adaptively tunes the consistency level at run-time. The main aim is to reduce the monetary cost (storage, network, and other related costs), along with a low fraction of stale reads.	Advantages: Cost Effective (Reduces Instances cost, Storage cost, and Network cost). Disadvantages: Average Consistency cost efficiency.

Table 3. Cont.

Replication Strategy Basic Details	Replication Strategy	Advantage and Disadvantages
[104] Year 2012	Description: An adaptive replication strategy that redeploys dynamically large-scale various file replicas on different data nodes and selects the data files which require replication based on minimal cost in order to improve the system availability.	Advantages: Cost effective, Low response time, Low bandwidth consumption, reduced waiting time, and High data access speeding up. Disadvantages: Less data availability.
[105] Year 2015	Description: A runtime-based replica consistency mechanism, namely (RBRC), which is mainly used for cloud storage systems. The mechanism achieves a dynamic balance between performance and consistency using read frequency. This method is based on access frequency and its access time.	Advantages: Decreased average file access time, Low replication delay time. Disadvantages: Average load balancing.
[106] Year 2015	Description: An adaptive consistency guarantee model that probes the consistency index of an observed replicated data object in an online application. The main aim is to reduce response time.	Advantages: Maintained response time and Time delay. Disadvantages: Assumed Load balancing setting for the Implementation.
[99] Year 2017	Description: A novel replication strategy which is used to reduce data storage cost in workflow applications. The strategy considers various parameters for the cost-related effectiveness, which include access frequency, data center storage capacity, the constraints of dataset dependency, and size of datasets in the build-time stage.	Advantages: Reduced cost of data management, decreased data movement, and decreased data transfer cost. Disadvantages: Increased response time.
[107] Year 2016	Description: A dynamic cost-aware replication strategy, which optimizes and identifies the least number of replicas that are required to maintain desired availability along with data reliability.	Advantages: Low replication cost, High reliability, and High availability. Disadvantages: Low consistency rates, Low load balancing, and High response time.
[108] Year 2013	Description: A response time-based replica strategy, namely RTRM, consisting of replica creation methods. The aim is to automatically increase the number of replicas based on average response time while maintaining the performance.	Advantages: High performance, Low response time, High rapid data download, Low energy consumption, and High data availability. Disadvantages: Low reliability, Low load balancing, and High replication cost.
[109] Year 2013	Description: A modified dynamic data replication strategy with synchronous and asynchronous updating. The work is based on the decision of a reasonable number of replicas, along with the right location of replicas, while keeping in mind the execution time.	Advantages: Execution time, High availability, and Performance. Disadvantages: Low speed data access and Low load balancing.
[110] Year 2014	Description: A dynamic replica selection and placement strategy which is used for cloud replica management. A replica creation is adapted continuously by changing network connectivity and users. It designs an algorithm for suitable optimal replica selection and placement with a target to increase data availability.	Advantages: Low access time, Low response time, low access cost, Shared bandwidth consumption, and delay time. Disadvantages: Low Load balancing.
[111] Year 2015	Description: An effective dynamic replica placement algorithm, namely BPRA, which is based on minimal blocking probability. The main intention is to improve the load balancing using user access information.	Advantages: Improved load balance, Reduced access skew, and file access latency. Disadvantages: Ignored QoS

Table 3. Cont.

Replication Strategy Basic Details	Replication Strategy	Advantage and Disadvantages
[52] Year 2019	Description: A data replication strategy for MongoDB. The main aim is to provide the performance requirement for the tenants, while the provider's profit is not ignored.	Advantages: Decreased response time, Resource consumption, and number of replications. Disadvantages: Low load balancing.
[112] Year 2018	Description: A predictive approach, namely (PredRep), which is used to characterize the cloud database system workload and automatically provide or reduce resources based on the cost factor and SLA agreement.	Advantages: Reduced cost and SLA violations. Disadvantages: Average load balancing.
[3] Year 2017	Description: A data replication strategy for cloud systems, namely (DPRS), which uses the number of requests and free storage space to determine the number of replicas along with a suitable placement site.	Advantages: Low response time, Enhanced storage space, and Effective network usage. Disadvantages: Low reliability.
[113] Year 2016	Description: A replica replacement strategy that considers the data file availability, the last time the replica was accessed, access number, and the replica size. The replication not only provides load balancing but also maintained the performance.	Advantages: Increased Performance and Load balancing, less storage usage. Disadvantages: Missing real time Implementation.
[114] Year 2018	Description: A dynamic adaptive replica strategy, namely (DARS), which uses node's overheating similarity to provide the replica creation time, the replica creation opportune moment and locate optimal replica placement node.	Advantages: Superior performance and Better load balance. Disadvantages: Lower access delay.
[115] Year 2020	Description: A data Replication Strategy (RSPC) that satisfies both performance and minimum availability tenant objectives while ensuring an economic profit for the provider in Cloud datacenters.	Advantages: Reduced resource consumption, Reduced Costs of provider (penalty and data transfer costs) Disadvantages: Missing real-time cloud implementation and consistency consideration.
[116] Year 2019	Description: A cost-based dynamic replication strategy (DRAPP) that uses the least number of replicas for simultaneous availability of data and performance tenant requirements in regard while considering the tenant budget along with a profit of provider. While dealing with tenant budget, query scheduling is done in such a way that replicas effectively obey load balancing.	Advantages: Reduced query response time and increased availability. Disadvantages: Missing real-time cloud implementation and energy consumption consideration.
[117] Year 2018	Description: A cost-based data replication strategy (PEPRv2) for cloud-based systems that effectively satisfies the response time objective (RTO) for executing queries while simultaneously benefiting the provider to return a profit from each execution. It simultaneously satisfies both the SLA terms and profit of the provider. The SLA includes the availability and performance along with maintaining the query load as per the provider's profit.	Advantages: Reduced response time, bandwidth consumption, and monetary expenditure. Disadvantages: Missing real-time cloud implementation.

In this section, we propose a taxonomy of target-oriented replication strategies based on target objective classification (shown in Figure 4). We classify these target-oriented replication strategies into nine key target objectives based on their attributes, namely (a) Objective 1: Availability, (b) Objective 2: Reliability, (c) Objective 3: Performance, (d)

Objective 4: Fault Tolerance, (e)Objective 5: Load Balance, (f) Objective 6: Scalability, (g) Objective 7: Elasticity, (h) Objective 8: Consistency, and (i) Objective 9: Cost.

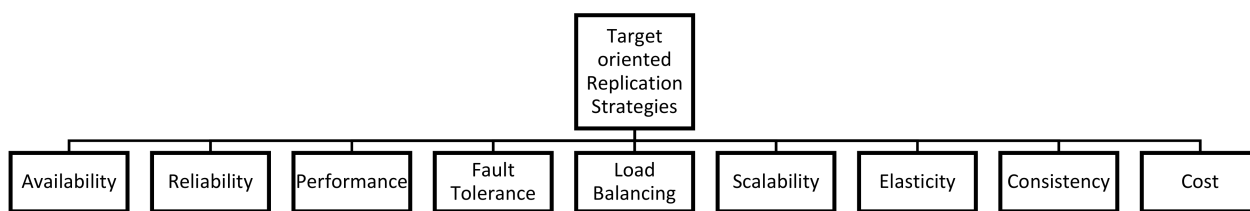


Figure 4. Demonstrates the classification of target-oriented replication strategies based on target objectives.

There is always a conflict between each targeted objective of replication strategies. For example, costs are inversely proportional to access time and performance. In fact, due to the different nature of each target objective, most of the replication strategies simultaneously do not satisfy the multiple target objectives. Each target-oriented replication strategy aims to satisfy a specific target objective to enhance the performance directly or indirectly. The first and crucial target objective always aims to improve the data availability, which is a must for accessibility and disaster recovery. The other important target objectives include increasing fault tolerance and throughput, providing reliability, scalability, elasticity, ensuring load balancing, decreasing response time, and security. In the future, a hybrid multi-objective replication approach can be planned and designed, like in Reference [73], which will possess the mixed capabilities of all target objectives.

4.6.2. Target Objectives of Target-Oriented Replication Strategies

In this section, different dynamic replication techniques related to their concern target objectives are explained in detail.

Availability

Availability is the readiness for correct service of a system [101] that guarantees that an item (data or service) is functioning at a given instance of time under defined conditions. Data availability has been always a hot topic and a big factor in the field of distributed environments that promises to improve the data or service available to the users for a better quality of service. Even if there are not any disasters, the data availability should be considered as the primary concern for the organizations for accessibility and smooth functioning. This is the reason why it is considered as a (Primary) main target objective for the replication strategies in the cloud. In all distributed database environments, and especially in cloud computing, the replication strategies target improving the availability of data. The replicating services always guarantee the availability of services in case of disasters. The large-scale distributed storage systems use replication strategies regularly to improve the data or service available to users. The two metrics which affect the data available in these types of setups are the number of replicas and the location of the replicas [118]. Availability is directly proportional to its reliability. There are many other metrics which affect data availability and must be addressed utmost. They include network link failures issues, replica allocation, and many more.

Reliability

Reliability [102] aims to give a correct or acceptable result within a time-bounded environment. Data reliability is an important concern in distributed environments. Many efforts have been used to improve the data reliability for the storage distributed environments. High reliability is always another main target objective for cloud storage systems. Replication strategies have multipurpose efficiency on data reliability and availability. As the number of replicas(availability) increases, there are more chances that the user's request will be serviced faster and hence more reliable will be the system. The metrics which affect the data reliability in distributed setups are the disk failure rate issues, number of replicas,

and response time, which keeps on increasing with an increasing number of tasks [31]. Many other metrics affect data reliability and must be addressed to the utmost. They include data missing rate, storage cost consumption, and effective data replica schemes for decent reliability [102]. Various research has been done in the respective field [119], and the work includes reliability issues of large-scale storage systems and provides a desirable solution for them.

Performance

Replication is an effective way to increase performance in a cloud computing environment by completing service requests from various users. Performance represents the effectiveness of the system [73]. The data storage must be in a strong condition to strengthen fast and strong data access, update management, and should provide recovery facilities. The performance in large-scale cloud storage systems is always considered as one the important topic and major target objectives to be addressed. Nevertheless, availability increases the performance of data in a distributed environment. Moreover, the replication strategies have multipurpose competence in data availability, data reliability, load balancing and response latency [120]. System performance must be achieved at an acceptable cost. The performance is computed in terms of throughput, response time, latency, and so on which also displays the quality of the service. The metrics which affect the data performance include: (1) Response time—time is taken by a system to respond to a service request, which should be low; (2) Throughput—number of service requests served at a given time, which should be high; (3) Latency—time delay of a client request and to its service providers response in the cloud; and (4) Execution time—service time to process the sequence of activities [73]. There are many other metrics that affect the performance and must be addressed at the utmost. They include the number of replicas which is directly proportional to availability and mostly enriches the performance.

Fault Tolerance

The stored data must have the option to recover if there is any occurrence or prediction of failures in one machine, which means the system should provide a backup instance of the application (data is still available on another machine on the network) that will commence or is expected to start without interruption [121]. Hence, fault tolerance techniques minimize the failure effect on the computing environment. Fault tolerance in cloud computing improves reliability, availability, recovery from failure, lower cost, improves performance metrics, etc. More chances of failures arise because of the dynamic behavior of cloud or distributed environments. To overcome such effects of these failures, the cloud should implement fault tolerance aggressively, which is always a crucial target objective to be considered while choosing or developing a replication strategy [122]. Replication increases the fault-tolerant by introducing a balance between consistency and performance during update scenarios. We need to have minimum latency for an efficient fault tolerance [121]. Hence, low latency (network delay), service time, and fewer overheads are the metrics of fault tolerance. Another metric can be the number of replicas, which needs to be in control to maintain the fault tolerance [123]. Fault tolerance provides resilience to the cloud-based replication strategies.

Load Balancing

Load balancing is one of the central target objectives for data replication in cloud computing. In a distributed system, load balancing is the process of distributing and balancing the dynamic local workload (memory capacity, delay, or network load) among various nodes (available replicas) to maintain resource utilization and achieve higher job response time [79]. Replication strategies show multipurpose efficiency on load balancing. It improves the overall performance of the system. It utilizes the available resources hence reduces the resource consumption. It also helps to implement fail-over, provide scalability and avoid the performance bottlenecks [79,124]. The metrics which affect the load balance

in this distributed computing include response time, request loss rate, optimal number of copies, and the storage [120,125].

Scalability

Scalability is another crucial target objective which needs to be addressed for optimal replication on cloud. Scalability is a capability of a system to handle the increasing demand for computational resources to accommodate the growth [90]. Scalability enhances the replication [126]. The requirements of cloud computing are scalability with large data set operations [90], resulting in increasing the performance using over-provisioning of the resources [127]. The data on storage systems needs prompt scale to cover the increasing workload demands by providing the provision to horizontal or vertical expansions [128]. Many of the cloud base applications rely upon data-replication to achieve better performance, availability, scalability, and reliability [129]. Elasticity is an extended version of scalability.

Elasticity

Elasticity is one more important target objective used to face the changing conditions during the replication of clouds. Elasticity is the capacity to expand or shrink, the number of replicas to adjust to the incoming increasing or decreasing workload [130]. Using elasticity, additional computational resources can be acquired, or released automatically (resources provisioned to their applications) based on demand (dynamic workload) to minimize the resource cost and filling the Quality of Service (QoS) requirements. Auto-scaling is another name for elasticity. However, overprovisioning causes resource wastage and extra monetary cost, while under-provisioning leads to performance degradation and violation of service-level agreement (SLA) [131]. So, while developing an elastic replication strategy, there should be utmost consideration on over-provisioning and under-provisioning circumstances.

Consistency

Consistency of replica placed is one of the important and crucial parameters which needs to be addressed for the optimal replication strategy on the cloud. Using data replication strategies, a data-intensive application can accomplish fault tolerance, improved availability, and data recovery [8]. There are many techniques used to enhance the consistency of replication on the cloud. In distributed systems (cloud), the data consistency is described as a mutual deal between data availability and partition tolerance in the CAP theorem (Brewer's theorem) [132]. The CAP theorem mentions that, out of three properties, the only properties can be accomplished at the same time inside an appropriate framework [132]. In this regard, the consistency alludes to the prerequisite that the clients should neither feel or be aware of working on a single node, nor should they be aware of the number of replicas used or assigned to them.

Cost

Cost is one of the important target objectives of replication strategies is the cost. The costs associated with replication strategies can be a storage cost or data transfer costs (Replication Cost) [115]. The preference must be given to the for economic reasons and for choosing a replication strategy. The cost of replication of a data file is different in different data centers and keeping in view the heterogeneous nature of the system, the cost of replication, availability, and performance should be contemplated together for optimal replication [107]. The metrics which affect the cost in cloud-based replication strategies include data moment, and cost of data transfer, dataset dependency, access frequency, storage capacities of data centers, and size of datasets in the build-time stage. The optimized data placement strategies can reduce the data movement and save data transfer costs among different data centers [99].

Various research has been done for the cost and its effective utilization in cloud systems; some include the electricity price-aware consideration [133], some include the replication cost-related efficiency, some of the works [134] address the storage space limitations, and some address general monetary costs of replication [135]. In recent times, the monetary costs if considered with tenant and providers profit had become a trend due to its nature of benefiting both parties (tenant and provider). These monetary-based replications strategies have been classified into provider-centric and consumer-centric strategies, both primarily focusing on the service providers profit and tenants' profit [115].

In all target-oriented replication strategies, QoS should integrate all the above-mentioned objectives (availability, reliability, performance, fault tolerance, load balancing, scalability, elasticity, consistency, and cost) to achieve the highest level of target achievement for optimal replication. The SLA contract represents the agreement between a service provider and its customers (agreed-upon guarantees) to guarantee assurance [6,136] to support the basic objectives like data availability, enhanced reliability, performance, etc. [137]. Furthermore, the service provider does not satisfy the performance levels due to the inherent network latency of the Internet. User expectation of QoS is always high, so it is mandatory to address the basic and architectural issue, in particular, what will happen and who is responsible, as well as set the tolerance level of business processes [90].

4.6.3. Target Objectives and Their Relationship with Parameters

Each replication strategy can address one or more target objectives and each targeted objective is composed of one or more attributes (parameters). Different replication strategies cover different parameters based on the target objectives. These attributes act as important metrics for the evaluation of the replication strategies in the cloud. In Table 4, various target objectives and their attributes are evaluated.

Table 4. Represents various target-oriented replication strategies with their target objectives based on their attributes, purpose, and metrics.

Replication Strategy	Target Objectives (Priority Based)	Attributes (Parameters)-Metrics
[100] Year 2011	<ol style="list-style-type: none"> 1. Reliability (Primary Target Objective) 2. Fault Tolerance 	<p>The storage space and the storage cost are the reliability-related attributes. Both attributes are reduced and are based on the need basis of replicas. It initially stores only one replica and is incremental in nature.</p> <p>The failure rates of storage units are the fault tolerance related attribute, which directly affects the fault tolerance. Hence, lower the probability of the failure, the higher will be the reliability.</p> <p>Note: The key investigating parameters are Data Reliability, Storage Space, Storage Cost, and Failure Rates.</p>
[101] Year 2012	<ol style="list-style-type: none"> 1. Availability (Primary Target Objective) 2. Performance (Response Time) 3. Load Balance 	<p>The replica no. is the availability related attribute, which is based on a mathematical model to maintain the number of replicas and availability requirement accordingly.</p> <p>The execution rate, response time and bandwidth consumption are the performance-related attributes, and they are reduced because of balanced replica placement.</p> <p>The replica placement is the load balance-related attribute, which is achieved by placing the most popular data files based on access history (access information of data centers).</p> <p>Note: The key investigating parameters are Data Availability, Number of Replicas, Response Time, Execution Rate, and Bandwidth.</p>

Table 4. Cont.

Replication Strategy	Target Objectives (Priority Based)	Attributes (Parameters)-Metrics
[102] Year 2015	<ol style="list-style-type: none"> 1. Reliability (Primary Target Objective) 2. Performance (Cost) 3. Fault Tolerance 	<p>The replica no. is the reliability-related attribute, which initially stores only one replica (original copy of the data).</p> <p>The storage space is the performance-related attribute, which is reduced and, hence, reduces the storage cost of the data. It provides cost-effective reliability based on cost and failure rates of storage units (fault tolerance).</p> <p>The failure rates of storage units are the fault tolerance related attribute. Fewer failure rates of storage increase reliability.</p> <p>Note: The key investigating parameters are Data Reliability, Number of Replicas, Disc Failure Rates, storage space, storage cost.</p>
[26] Year 2012	<ol style="list-style-type: none"> 1. Performance (Primary Target Objective) 2. Availability 	<p>The stale reads rate is the performance-related attribute, which defines the consistency requirements of the application and affects the performance.</p> <p>The replica no. is the availability related attribute, and it dynamically adjusts the number of the replicas used in operation according to the run time based estimated stale read rate and network latency.</p> <p>Note: The key investigating parameters are Data Availability, Number of Replicas, and Stale Read Rates.</p>
[103] Year 2013	<ol style="list-style-type: none"> 1. Consistency (Primary Target Objective) 2. Performance (Cost) 	<p>The stale reads rate and relative cost of the application are the consistency-related attributes, which are estimated based on a probabilistic model using the current read/write rate and network latency. Stale reads are the output of access patterns exhibited by the applications. A low fraction of stale reads is maintained.</p> <p>The consistency cost is the performance-related attribute, consistency is chosen based on operations and is presented by the number of replicas in the quorum (a subset of all the replicas).</p> <p>Note: The key investigating parameters are Stale Reads, Consistency, and Cost.</p>
[104] Year 2012	<ol style="list-style-type: none"> 1. Reliability (Primary Target Objective) 2. Availability 3. Performance (Cost) 	<p>The replica no. is the reliability-related attribute. It improves the data reliability of files based on prediction of the past data access user requests using (Holt's Linear and Exponential Smoothing (HLES)) time series technique.</p> <p>The optimal replica no. is the availability related attribute. It chooses the best optimal replica selection and placement for the availability purpose.</p> <p>Low replication cost and average response time are the performance-related attributes.</p> <p>It particularly minimizes the bandwidth consumption of the data and increase the load balancing.</p> <p>Note: The key investigating parameters are Data Availability, Number of Replicas, Bandwidth, and Load Balancing.</p>
[105] Year 2015	<ol style="list-style-type: none"> 1. Consistency (Primary Target Objective) 2. Performance 	<p>Replica read frequency is the consistency related attribute. The replicas with high read frequency are updated aggressively, and low read frequency replicas are updated in a lazy way.</p> <p>The other parameters include average file access time, percentage of requesting up-to-date data and number of replications.</p> <p>File access delay time is the performance-related attribute. It lowers the number of replications without wasting network bandwidth, and, because of its shorter replication time, the file access delay time is also reduced.</p> <p>Note: The key investigating parameters are Replica Read Frequency, no. of Replicas, and File access delay time.</p>

Table 4. Cont.

Replication Strategy	Target Objectives (Priority Based)	Attributes (Parameters)-Metrics
[106] Year 2015	1. Consistency (Primary Target Objective) 2. Performance	Time gap and consistency tuner (consistency index-based protocol-which is the number of correct reads over the total reads are the performance-related attributes. The other parameters include the number of replicas and the threshold of a time gap, which is a minimum value of time gap between a succeeding read request and an update. Note: The key investigating parameters are Time Gap and no. of Reads.
[99] Year 2017	1. Cost (Primary Target Objective)	The storage cost is the performance-related attribute. The other parameters which play an important role and affect the performance directly include access frequency, storage usage or capacity of data centers, dataset size, and data dependency. Note: The key investigating parameters are Performance, Cost, and Data Size.
[107] Year 2016	1. Cost (Primary Target Objective) 2. Availability	The storage cost is the performance-related attribute, which is based on the least number of replicas required for a proper availability, the data file is selected on the basis of access intensity, the higher SBER, better response time, and cost of replication. The system byte effective rate, bandwidth consumption, and the response time are the availability related attributes. Note: The key investigating parameters are data file availability, Average File Probability, cost of replication, data file availability, system byte effective rate, and the cost of the replication.
[108] Year 2013	1. Performance (Primary Target Objective)	The response time is the performance-related attribute. When the response time is longer than the threshold, the replica number will increase; hence, the system will create a new replica. In addition, other related attributes are network utilization, average job time high rapid download and low energy consumption. Based on the new request, the bandwidth is predicted for replica selection. Note: The key investigating parameters are replica creation, Replica selection, and Replica placement.
[109] Year 2013	1. Availability (Primary Target Objective) 2. Performance	The replica no. is the availability related attribute. The number of replicas is considered as system byte effective rate and is calculated as the number of bytes available to total bytes requested by all tasks. The system byte effective rate is performed in the second stage of the Modified D2RS algorithm stage which is best suited for varied periods. The execution time is the performance related attribute, which increases the performance. Execution time is increased by creating a replica of the data in the data center. The popularity degree is the access frequency based on time factor and user activity. Note: The key investigating parameters are Data Availability, Number of Replicas, Execution Time, and Access Frequency.
[110] Year 2014	1. Availability (Primary Target Objective)	The replica no. is the availability-related attribute, which is based on the demands of the users and the availability of storage. It chooses the optimal replica selection and placement for the availability purpose based on response time and access time. Note: The key investigating parameters are Data Availability, Access Time, and Response Time.
[111] Year 2015	1. Reliability (Primary Target Objective)	The replica placement is the reliability-related attribute, which improves reliability and reduces access skew. The reliability is achieved through access latency (decreased file access latency). Note: The key investigating parameters are Data Availability, Access Latency, and Replica Placement.

Table 4. Cont.

Replication Strategy	Target Objectives (Priority Based)	Attributes (Parameters)-Metrics
[52] Year 2019	<ol style="list-style-type: none"> 1. Performance (Cost) (Primary Target Objective) 2. Load balancing 3. Fault tolerance 	<p>The response time of query and no. of replicas are the performance-related attributes. The former includes the data size, number of shards, I/O, and network bandwidth. The latter is responsible for data placement based on the estimated threshold, access frequency, and response time.</p> <p>The network bandwidth and least resource consumption reduce the communication costs respectively.</p> <p>The high access frequency is load balancing related attribute, which selects only the most popular data for replication.</p> <p>The sharding is the process of parallelizing the data by splitting the data uniformly across clusters. Hence, sharding is the fault tolerance related attribute.</p> <p>Note: The key investigating parameters are Data Reliability, Storage Space, Storage Cost, and Failure Rates.</p>
[112] Year 2018	<ol style="list-style-type: none"> 1. Elasticity (Primary Target Objective) 2. SLA (Service Level Agreement) 	<p>The variation of the response time and satisfaction function/SLA is the elasticity related attributes. The former is based on the SLA metric which is the value of maximum time response, and the latter is directly based on SLA compliance.</p> <p>Note: The key investigating parameters are Tenant size and Databases (size, queries, and response time).</p>
[3] Year 2017	<ol style="list-style-type: none"> 1. Performance (Cost) (Primary Target Objective) 2. Fault Tolerance 	<p>The number of replicas is the performance-related attribute, which is minimized and the appropriate sites for data placement are dependent on the number of used requests, site centrality, and storage. The knapsack provides the cost optimization of the replication.</p> <p>The other performance-related attribute includes better response time and less cost of replication, decrease user-waiting time, and improve data access.</p> <p>Data access popularity and parallel download are the fault tolerance related attributes. The effective network usage mean response time, storage usage, replication frequency, and hit ratio with respect to the others are also considered as an enhancement achievement.</p> <p>Note: The key investigating parameters are Response time, Storage usage Effective network usage, Replication frequency, and Hit ratio.</p>
[113] Year 2016	<ol style="list-style-type: none"> 1. Performance (Cost) (Primary Target Objective) 2. Load balancing 	<p>The replacement strategy is the performance-related attribute, which is based on the availability of the file, the last time the replica was requested, the number of access, and the size of the replica.</p> <p>Other performance-related attributes include cost, which relies on as storage size of each site, which is kept limited by just keeping the important data only.</p> <p>The replica placement policy is the load balancing related attribute, which allows storing replicas in the relevant sites based on five parameters (failure probability, storage usage, mean service time, latency, and load variance).</p> <p>Both Performance and Load Balancing related attribute target to increase the response time and cost-effective availability.</p> <p>Note: The key investigating parameters are mean Response time, Load balancing, Effective network usage, Replication frequency, and Storage usage.</p>

Table 4. Cont.

Replication Strategy	Target Objectives (Priority Based)	Attributes (Parameters)-Metrics
[114] Year 2018	1. Performance (Primary Target Objective) 2. Load balancing	<p>The replica creation time and opportune moment are the performance related attribute, which is based on the node's overheating similarity. They find the optimal placement node using the fuzzy clustering analysis method, and then the replicas are created by node using a decentralized self-adaptive manner. The optimal placement node is the load balancing related attribute, which is found from the neighborhood. The optimal placement node improves the probability of replica to be accessed, relieves the overloaded high node degree, possess low node load, reduces the access delay, and boosts the load balance.</p> <p>Low access delay and acceptable load balance are achieved by reducing the node response latency. Hence, low access delay is based on operation time and the ratio of request versus response. Note: The key investigating parameters are Low access delay, Ratio of average load, Node response latency, and Accessing pressure.</p>
[115] Year 2019	1. Cost (Primary Target Objective)	<p>The response time of the query is the cost-related attribute, which is responsible for data placement based on the critical threshold achieved. The replica factor is dynamically adjusted to reduce resource consumption. Replica creation relies on the minimum availability objective and Response time (RT) objective. The Strategy always keeps the minimum number of replicas. RSPC satisfies the response time requirement under high loads, complex queries, and strict response time thresholds. Note: The key investigating parameters are Response Time and Replication Cost.</p>
[116] Year 2019	1. Cost (Primary Target Objective)	<p>The number of replicas is the cost-related attribute, which is responsible for effective load balancing based on query scheduling. The other replication attribute includes response time and SLA agreements. The former is dependent on the threshold. The availability should be less, or response time should be greater than a threshold for effective replication, and the latter minimizes the SLA violations. Note: The key investigating parameters are Load balancing, Response Time, Bandwidth Consumption, and Cost.</p>
[117] Year 2018	1. Cost (Primary Target Objective)	<p>The response time of the query is the cost-related attribute. The execution of any particular query is estimated and compared with service level objectives (service quality) that the tenant expects from the provider along with profit estimation. It also decreased the number of replicas for a given availability. Note: The key investigating parameters are Response Time, storage usage Network bandwidth consumption, and Cost.</p>

4.6.4. Quantitative Analysis of Target-Oriented Replication Strategies

Data replication in distributed file systems (clouds) is a technique to store the data (replicas) on multiple servers across multiple data centers with the main aim to improve data availability during failures. The other advantages of replication are to improve the response time, bandwidth consumption, reliability, job performance, throughput, less frequency, reduce data access latency, decrease data transfer amounts, and the costs [138,139].

The focus of each target-oriented replication strategy is to satisfy a specific target objective following its prescribed matrices to increase the overall performance. We have observed that several dynamic replication strategies discussed in this article trend to address most of the primary addressed target objects (most addressed). Some of the dynamic replication strategies address secondary target objectives (average addressed), and a few of the dynamic replication strategies address tertiary target objectives (least addressed).

From Table 5, we have observed that some of the strategies, like Refs. [3,52], are included in the fault tolerance category and are also included in the performance category; the same had happened with Refs. [100,102], which are included in both fault tolerance and reliability section. There are other, same examples in Refs. [52,101,113,114], found in the performance and availability category, and also Reference [112] is included in scalability, elasticity, and cost category because all these strategies are addressing primary and secondary target objectives in a single replication strategy, with each target objective having its priority. In other words, these strategies address both categories of target objectives simultaneously.

Table 5. Summary of all target objectives (Most addressed target objective, Average target objective, and Least target objective).

Most Addressed Target Objective		Average Addressed Target Objective		Least Addressed Target Objective	
Availability	[101,109,110]	Fault Tolerance	[3,52,100,102]	Scalability	[112]
Reliability	[100,102,104,108,111]	Load Balance	[52,101,113,114]	Elasticity	[112]
Performance	[3,26,52,108,113,114]			Consistency	[103,105,106]
				Cost	[99,107,115–117]

5. Performance Evaluation of Target-Oriented Replication Strategies: Comparison and Evaluation

Here, we provide a complete and detailed survey for the target-oriented replication strategies in cloud with their attribute status and explanation, as depicted in Table 6.

5.1. Features of Target Objectives for Target-Oriented Replication Strategies in Cloud

Table 7 shows a summarized form of features included in all target-oriented replication strategies. In this section, we compare and evaluate the reviewed target-oriented replication strategies according to their features. These features are represented, along with their intensities, as LW for Low, MD for Medium, HG for High, IN for increased, NA for not addressed, YS for yes addressed, and NC for No Change.

Table 6. Comparison evaluation of different target-oriented replication strategies in cloud.

Attribute		Target-Oriented Replication Strategies for Cloud Computing										
		[3]	[26]	[52]	[102]	[100]	[101]	[103]	[104]	[105]	[106]	[99]
Availability	Achieved by decreasing replication frequency and maintaining the rational replicas during changing workloads		Increased	Achieved	Increased	Not Addressed	Increased	Increased by maintaining a low fraction of stale reads	Increased	Addressed	Increased	Not Addressed
Reliability		Not Addressed	Not Addressed	Not Addressed	Increased with least replication using proactive replica checking	Increased by predicting and dynamically generating an additional replica when needed	Not Addressed	Not Addressed	Increased by the lightweight time-series prediction algorithm	Not Addressed	Not Addressed	Not Addressed
Storage Space	Reduced by breaking the data file into different parts for best storage		Not Addressed	Maintained the storage through popularity degree based on least popular data files are removed and most popular accessed data are replicated	Reduced (limited)	Reduced	Not Addressed	Reduced	Not Addressed	Not Addressed	Not Addressed	Not Addressed
Storage Cost	Achieved through least resource utilization		Not Addressed	Reduced because it places the new replicas closer to data consumers and hence reduce the communication costs	Reduced	Reduced	Not addressed	Reduced	Reduced	Not addressed	Not addressed	Reduced due to parameter changes in data set dependency, access frequency, and by partitioning storage space

Table 6. *Cont.*

Attribute		Target-Oriented Replication Strategies for Cloud Computing									
	[3]	[26]	[52]	[102]	[100]	[101]	[103]	[104]	[105]	[106]	[99]
Bandwidth Consumption	Reduced by maintaining the rational replicas during the change environment sessions. Hence, decreases the unnecessary replication which effects directly to bandwidth	Not Addressed	Reduced because of removing the unnecessary replications	No Reduction	Not Addressed	Reduced due to balanced placement of replicas	Not Addressed	Reduced	Reduced because low read frequency replicas are updated in a lazy way	Not addressed	Not Addressed
Optimal Number of Replicas	Achieved because it determines the no. of replicas and suitable sites for replica placement based on the number of free spaces, requests and site centrality	Achieved	Maintained because the number of replicas is adjusted dynamically to reduce the resource consumption	Achieved	Not Addressed	Achieved	Achieved	Achieved	Achieved	Achieved	Not Addressed

Table 6. *Cont.*

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[3]	[26]	[52]	[102]	[100]	[101]	[103]	[104]	[105]	[106]	[99]
Response Time	Reduced by decreasing the user-waiting time	Reduced by using both the application requirements and the storage system state to handle the consistency at run time and also by using the stale read rate of the application	Response time is decreased, and its estimation is based on parameters that impact the query execution and threshold criteria	No Reduction	No Reduction	Reduced due to placement of replicas	Reduced due to maintenance of an acceptable rate of fresh reads	Reduced	Reduced due to higher percentage of reads, latest data average file access and delay time	No Reduction	Increased response time because of high network usage
Load Balancing	Achieved due to reduced replication frequency	Achieved through stale reads estimation	Maintained because only popular data, i.e., having a high access frequency are replicated	High	No Load Balancing	Achieved by placing the Replicas based on access history of data nodes	Achieved by selecting the highest consistency-cost efficiency level to adapt to workload dynamically	Achieved by placing replicas based on Heuristic search Algorithm	Not Addressed	Achieved by using round-robin policy, nearest replica, or heuristic-based replica selection	Not Addressed
Fault Tolerance	Achieved by using data access popularity	Not Addressed	Achieved by splitting the data uniformly across various clusters	Achieved	Achieved	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed

Table 6. Cont.

Attribute											
Target-Oriented Replication Strategies for Cloud Computing											
	[3]	[26]	[52]	[102]	[100]	[101]	[103]	[104]	[105]	[106]	[99]
Consistency	Future Work	Increased	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Increased	Not Addressed	Increased due to replica read frequency	Increased due to time gap effect and consistency index factor	Not Addressed
Scalability	Not Addressed	Increased	Scalability is achieved through auto-sharding	Not Addressed	Not Addressed	Not Addressed	Increased	Not Addressed	Increased	Increased	Not Addressed
Elasticity	Not Addressed	Increased	Achieved because it removes all unrequired resources/replicas	Not Addressed	Not Addressed	Not Addressed	Increased	Not Addressed	Not Addressed	Not Addressed	Not Addressed
SLA	Achieved as QoS requirements are fulfilled	Not Addressed	Achieved and gets triggered only when the response time of a tenant query is more than a response time threshold	Not Addressed	Not Addressed	Not Addressed	Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Availability	Maintained	Not Addressed	Increased due to replication of more recently accessed file	Increased	Increased	Achieved	Increased due to addressing of five parameters in load balancing strategy	Not Addressed	Addressed	Maintain minimum number of replicas for high availability along with performance	Achieved by maintaining minimum availability level

Table 6. Cont.

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Reliability	Maintained	Not Addressed	Not Addressed	Not Addressed	Increased due to decreased file access latency	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed
Storage Space	Reduced by deleting the additional replicas long term unassessed data files, or having least access rate compared to other data files	Not Addressed	Not Addressed	Not Addressed	Reduced due to reduced storage redundancy	Reduced by reducing the tenant storage size	Reduced due to the placement of data files in data nodes with low storage utilization to minimize the waiting time	Not Addressed	Reduced search space due to less creation of replicas	Reduced search area because of the known local budget of sub-region which determines the number of replicas	Reduced because from a selected subregion, a node with acceptable storage space is selected for a placement
Storage Cost	Reduced because knapsack algorithm is used to invoke to optimize the cost of replication	Not Addressed	Not Addressed	Not Addressed	Reduced due to replica factor, hence reduce the cost of data management	Reduced	Reduced because replicas are dynamically created in advance. Files can be stored in a specific, hence reducing the storage usage (storage elements usage (SEU))	Not Addressed	Reduced penalty and data transfer costs because most replications are performed per set of queries and also uses fewer storage resources due to fewer replicas creation	Reduced because replicas are only created if the providers profit goes higher than the replication cost	Reduced because it is estimated as the cost of storage I/O performed by any particular query

Table 6. Cont.

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Bandwidth Consumption	Reduced	Reduced because it predicts the bandwidth among the replica servers	Not Addressed	Reduced because during selection and placement of replica, it needs minimum bandwidth consumption	Maintained the bandwidth consumption which does not exceed the bandwidth despite various file request arrival data nodes	Reduced by provisioning	Reduced to reduce the reduced latency	Reduced because of increasing the service node overload	Reduced bandwidth consumption due to Network Bandwidth (NB) locality i.e., a replica of a required remote data is placed at a node having a larger NB toward the node requiring remote data	Reduced bandwidth consumption because the replicas are closer, which target less data transfer	Reduced due to better replica placement by evaluating each subregion for the profit satisfaction and response time and comparison
Optimal Number of Replicas	Achieved	Achieved because if the response time is longer than the threshold, the new replicas are created	Achieved	Achieved because optimal replica selection and replica placement rely on response time and access time	Maintained because through BPRA algorithm, the minimal number of replicas are calculated according to available requirement and also replica factor is dynamically adjusted based on file access frequency	Achieved based on forecast approach. Additionally, It provides the estimated time to create a replica, the replica size and information on each tenant	Maintained through data placement	Uses the fuzzy set model for the optimal node	Due to the majority of in region replicas creation, lesser number of replicas are generated, and a remarkable data transfer saving is achieved	Not addressed because search area is reduced	Achieved as it is always associated with each query being executed at a particular time and follow the near-optimal placement

Table 6. Cont.

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Response Time	Reduced even after replication cost equals to the budget	Reduced because during concurrent requests to file, it maintains the average service response time	Reduced	Reduced because response time rely on bandwidth utilization of each random file	Maintained	Reduced	Reduced through increasing the total number of local accesses and avoiding the unnecessary replication	Reduced because of lowest access delay (lowest nonresponse ratio)	Satisfies the response time requirement under high loads (Acceptable Response Time) along with providers profit, especially during high loads. Average response time is reduced due to less Virtual machines overload	Reduced because it uses query number and if providers gain is real, then only a new replica is created	Reduced and checked if it satisfies the service quality expectations from the tenants
Load Balancing	Not Addressed	Achieved by maintaining replica placement location			Achieved because it selects the optimal data node with the minimal blocking probability	Maintained	Achieved through the data placement	Increased due to fuzzy clustering analysis method used to select the optimal placement node for stored replicas	Achieved because it does not replicate data when response time objective is satisfied	Achieved because data replication and scheduling of queries are coupled using tenant budget using node load criterion	Reduced because, from a selected subregion, a node with an acceptable load is chosen for the data placement

Table 6. Cont.

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Fault Tolerance	Not Addressed	Not Addressed	Not Addressed	Achieved automated fault tolerance	Not Addressed	Achieved by using write-ahead logging scheme	Achieved by improved response time, which generates the new replicas and storing them to the less-load sites	Not Addressed	Not Addressed	Not Addressed	Not Addressed
Consistency	Future Work	Not Addressed	Not Addressed	Not Addressed	Not Addressed		Not Addressed	Future Work	Not Addressed	Not Addressed	Not Addressed
Scalability	Not Addressed	Not Addressed	Not Addressed	Achieved because the utility computing is based on scaling and scaling is based on optimal replica selection	Not Addressed	Achieved by using Scale tenant	Not Addressed	Increased because of a self-adaptive feature which can handle overload on time and reduces the access delay and hence increase the scalability	Not Addressed	Not Addressed	Not Addressed

Table 6. Cont.

Attribute	Target-Oriented Replication Strategies for Cloud Computing										
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[125]	[126]	[127]
Elasticity	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Achieved by using the predictive elastic replication strategy which utilizes process decision, based on satisfaction function and the variation of response time	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed
SLA	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Not Addressed	Achieved by using the forecast service to assist the decrease of SLA violations.	Not Addressed	Not Addressed	Maintained, addressing both tenant and providers benefits	Due to its dynamic nature, the number of SLA violations are low (Very Less)	Maintained while considering SLA violation count as a measure of response time satisfaction

Table 7. Features of target objectives for target-oriented replication strategies in cloud.

	[3]	[26]	[52]	[102]	[100]	[101]	[103]	[104]	[105]	[106]	[99]
Availability	YS	IN	YS	IN	NA	IN	IN	IN	YES	IN	NA
Response Time	LW	LW	LW	NC	NA	LW	LW	LW	LW	NC	IN
Reliability	NA	NA	NA	HG	HG	NA	NA	HG	NA	NA	NA
Bandwidth Consumption	LW	NA	LW	NC	NA	LW	NA	LW	LW	NA	NA
Load Balancing	YS	HG	YS	HG	NA	HG	HG	HG	NA	HG	NA
Storage Cost	YS	NA	LW	LW	LW	NA	LW	LW	NA	NA	LW
Consistency	YS	HG	LW	NA	NA	NA	HG	NA	HG	HG	NA
Fault Tolerance	YS	NA	YS	YS	YS	NA	NA	NA	NA	NA	NA
Optimal no. of replicas	YS	YS	YS	YS	NA	YS	YS	NA	YS	YS	NA
	[107]	[108]	[109]	[110]	[111]	[112]	[113]	[114]	[115]	[116]	[117]
Availability	NC	NA	IN	IN	IN	YS	IN	NA	YS	YS	YS
Response Time	LW	LW	LW	LW	NC	LW	LW	LW	LW	LW	LW
Reliability	NC	NA	NA	NA	IN	NA	NA	NA	NA	NA	NA
Bandwidth Consumption	LW	LW	NA	LW	NC	LW	LW	LW	LW	LW	LW
Load Balancing	NA	YS	NA	NA	YS	YS	YS	IN	YS	YS	LW
Storage Cost	LW	NA	NA		LW	LW	LW	NA	LW	LW	LW
Consistency	NA	NA	NA	NA	NA	LW	YS	NA	NA	NA	NA
Fault Tolerance	NA	NA	NA	YS	NA	YS	YS	NA	NA	NA	NA
Optimal no. of replicas	YS	YS	YS	YS	NC	YS	YS	YS	YS	NA	YS

LW for Low, MD for Medium, HG for High, IN for increased, A is for not addressed, YS for yes addressed, and NC for No Change.

5.2. Performance Evaluation Understanding

In our research, we included a total of 22 different target-oriented replication strategies (2011 to 2019) in cloud domain (shown in Figure 5), and each strategy is addressing a specific target objective, or several, by either addressing the one attribute or many attributes. We have observed that primary addressed target objects (most addressed target objective), which include availability, reliability, and performance, are covering total of 80 percent, and rest of the target objectives, which include the secondary target objectives (average addressed target objective), including fault tolerance, and load balancing and the tertiary objectives (least addressed target objective), consists of scalability, elasticity, consistency, and cost. covers the rest 20 percent. The elasticity covers the 5%, and the consistency covers the 15%. The others target objectives, like fault tolerance, load balance, scalability, and cost are addressed indirectly, along with directly addressed target objectives.

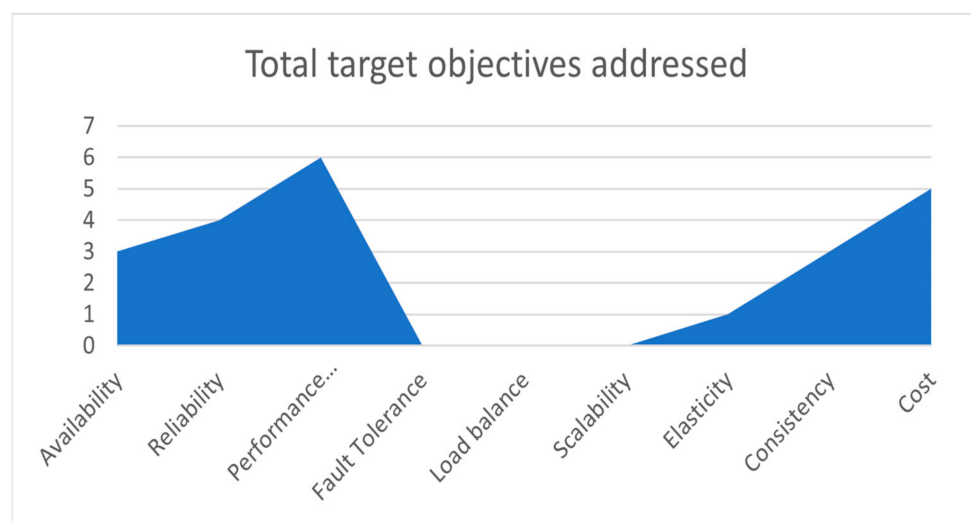


Figure 5. Total target objectives addressed in this research.

In future research, we recommend that least target objectives should be addressed with primary target objectives in a single replication, e.g., scalability should be considered with availability. Moreover, efforts should be made to develop a dynamic replication strategy that should address almost all (most addressed, average addressed, and least addressed) target the objective, altogether, in one algorithm. The detailed overview of all strategies included in this research paper is represented through Figure 6 (pie chart of quantitative analysis of target objectives).

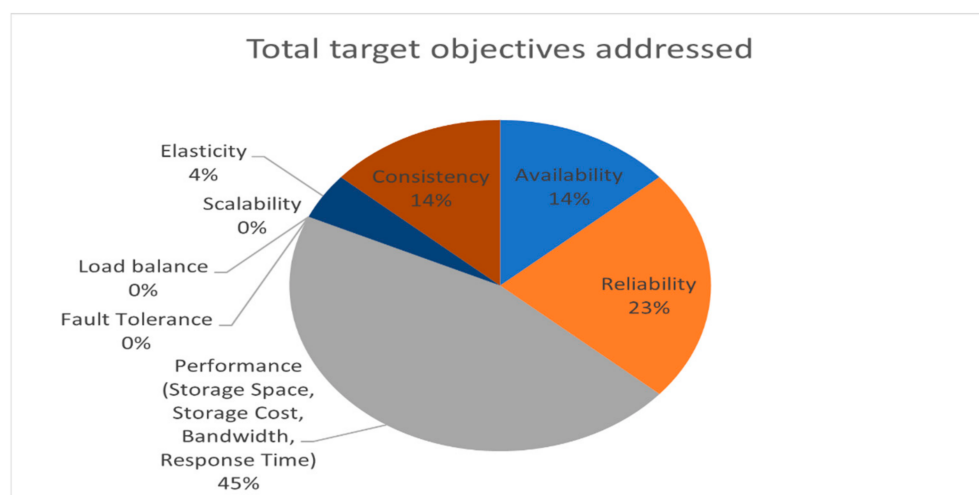


Figure 6. Pie chart of target objectives addressed directly in target-oriented replication strategies.

The functional metrics included in this work are the previously used performance metrics of cloud data replication and management for cloud systems [7]. Indeed, for the best optimization, the metrics discussed should contribute to increasing the overall performance by addressing many parameters of target objectives. The prime target includes the system availability, which is always a key factor for the overall enhancement and optimization. For a better system availability, the frequently accessed data is distributed to multiple suitable locations, from which the users can access the data from a nearby site [140].

In the future, these so-called metrics or the target objectives of target-oriented replication strategies in cloud computing strategies should also contribute to improving the security of the dynamic replication strategies, like in Ref. [141], because, indirectly, the security can lead to a data loss situation.

6. Challenges for Replication Strategies in Clouds

The main issues of replication revolve between data availability, cost, and performance. The frequently used data should get replicated to multiple locations to increase the data availability and enhance the performance; this will make a smooth way for the users to accessing from their nearby sites [54]. The other issues and challenges include data consistency, downtime during new replica creation, maintenance overhead, and lower performance [34].

Some of the latest work in the field of replication includes Refs. [142,143].

6.1. Challenges of Dynamic Replication Strategies in Clouds

Cloud replication primarily aims to increase the resource availability, reduce the delay time, minimize the access cost, and share the bandwidth consumption. During dynamic replication, decisions are made based on the resource availability and current access patterns.

There are two major issues in a replication which include: which data to replicate (replica selection) and where to place (replicas placement) [144]. Besides these two major issues in replication, there are also two other related issues, such as when to replicate (Replica time) and how many numbers of replicas to replicate (replica quantity). These other issues are as important as that of two major issues in replication [145]. Hence, the total of four important issues of any data replication strategy is determined as (1) what data should be replicated, (2) where to place a new replica, (3) when a replica should be created or deleted, and (4) how many replicas to create [52].

Some of the latest work in the field of dynamic replication includes Refs. [146,147].

6.1.1. Replica Selection

One of the major issues in cloud-based replications is replica selection. To meet the user requirement, such as to reduce the waiting time and increase the data access, replica selection must be addressed in cloud replications effectively. In adverse conditions, if early replication of a data file is done, or if the replica selection is not done efficiently, both conditions will lead unnecessary utilizing an extra storage space consumption and will increase its associated storage cost.

The available solutions can be the selection of a particular popular data [148] or the selection of data having the relatively higher reliability and longer storage duration, or we can try any light-weight time series prediction technique [31] to overcome the hurdles.

6.1.2. Replica Placement

One more vital issue in cloud-based replications is replica placement. The replica factor is one of the key factors of replica placement. Replica placement promotes data availability and service quality. The main two issues in the replica placement are how to determine the replica factor and how to select the optimal data node for storage of replica. Replica placement algorithm are categorized into two basic types: as static replica placement algorithm and dynamic replica placement algorithm. Static replica placement algorithm generates replica and selects data node at the initialization of the cloud storage system. These algorithms are easy to deploy, while dynamic replica algorithm selects the optimal data node dynamically to store replica based on current available data. These algorithms cannot be easily deployed [111,149].

To decide and address the issue of where to place the replica is a vital part and a crucial point in cloud computing architectures. As a solution, we can aim and stress on file access history used, which is a readily available solution.

The available solution for replica placement issues of different replication strategies uses blocking probability technique, which is used by paper [84], access information of data node technique used by reference [101], and heuristic search algorithm used by Refs. [31,104]. In general, we need to determine the best location and reduce the access latency factor for efficient replica management [150].

Many related surveys and open research issues are mentioned in Ref. [151], and some solutions for replica selection and replica placement in cloud setups are mentioned in Refs. [140,148,152].

6.1.3. Replica Time

Another important crucial issue in cloud-based replications to be addressed is when to replicate. The selection of proper time not only enhances the availability but also reduces the cost of the storage indirectly. The replica selection and replica time should correlate for an efficient output which includes the data availability, low-cost storage, and reliability. For an efficient cloud computing-based replication, this factor must be addressed at the utmost.

The available solution used in many cloud-based replications are mostly based on threshold achievement. These solutions can be either (1) the right time to replicate data is when the access frequency is greater the threshold, or (2) the right time to replicate data is when replica creation time point is reached, or (3) the right time to replicate data is when popularity exceeds the threshold, or (4) the right time to replicate data is when the original copy does not meet the user-specified reliability requirement, or (5) the right time to replicate data is when the replication factor is less than the specified threshold [145].

6.1.4. Replica Quantity

One more crucial factor in cloud-based replications is the replica quantity. Besides meeting the system availability, reliability requirement, and the cost of replica maintenance, one of the important issues to be addressed in cloud replications is how many numbers of replicas to replicate because, after a certain period of time, increasing the number of replicas does not increase the availability but might bring the unnecessary consumption of storage space, hence increasing the cost of storage. According to papers [84,150], it is very important to decide the replica quantity for cost-effectiveness purposes.

The available solutions used include mathematical model (built on the concept of theory of temporal locality, which states that there is a probability in the future that most recently accessed data file will be accessed again) to capture the relationship between the availability requirement and the number of replicas; another solution includes storage duration, the number of replicas and user-specified reliability requirement, and a few include the numbers of replicas to be calculated by a parameter smoothening factor [106,145].

Figure 7 depicts the issues and future research directions of dynamic replication strategies in clouds.

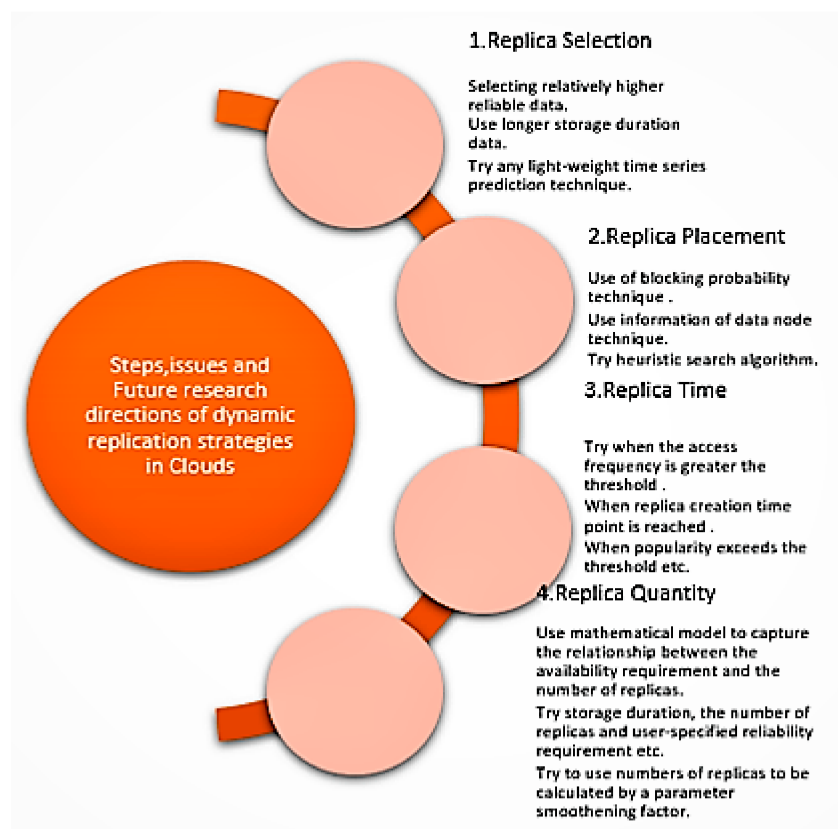


Figure 7. Issues and future research directions of dynamic replication strategies in clouds.

7. Least Addressed Target Objective of Target-Oriented Replication Strategies in Clouds, Their Challenges, Issues, and Future Research Directions

In this section, we discuss challenges, issues, and future research directions of tertiary addressed target objective one-by-one: Tertiary addressed target objective are the target objectives which are least addressed in all replication strategies. They are mentioned below as:

7.1. Scalability: Challenges and Issues

Due to the huge scale of the data stored in the data centers, there is always a need for quick scale to meet the workload demands. These huge data centers on a distributed setup are more prone to failures. Therefore, distributed cloud resources need to be efficiently utilized to minimize the costs associated with the storage and to maintain communication of these applications effectively along with the data availability. The replica locations and the associated communication cost are always a big concern for the replication strategies on the cloud computing paradigms.

For smoothness of storage, the cost-effectiveness and accommodating load spikes are considered as a big challenge. Furthermore, resource utilization must be adaptive for the flexibility of resource availability, for the flexibility to the addition of new resources, for the flexibility in case of load variations and for the distribution of client locations [66,153].

Hence, scalability is always considered an important metric that must be stressfully addressed by all replication algorithms. There are various factors which effects the scalability. The most important factor includes the architectural to be chosen. The architecture to be chosen for the replication plays an important role for the success of data replication. However, different architectural models (grid or cloud or other) possess various levels of scalability, which means that scalability is more dependent on the model, rather than replication algorithm.

Future Research Directions for Scalability

An analytical study shows that scalability depends more on the architecture model (grid or cloud or other) rather than the replication algorithm [66]. Different architectural models support different levels of scalability. Therefore, while targeting performance through scalability, we need to choose the architectural models of the cloud for the replication strategies.

Another way to improve the scalability of replication strategies on cloud is to use asymmetric processing, in which the transactions are initially processed at the originating location sites and then are collectively and eventually propagated to other sites, while, in symmetric processing, the updates are sent and executed at all replicated sites [154]. Some of the latest work in scalability include Refs. [155,156].

7.2. Elasticity: Challenges and Issues

Elasticity is the capability to expand (scale up) and shrink (scale down) the number of replicas according to incoming load. The resource provisioning issue is one of the biggest issues in distributed computing configurations [157], especially when we talk about the dynamic workload and dynamic environments. The available solutions include proactive and reactive approaches [112]. During high workloads, the data storage must be able to expand with increasing load hike and also adjust to shrink during low load by releasing the unutilized cloud resources [128]. Elasticity and scalability objective are two interrelated terms, where the latter allows the shrinking concept of the resources besides the expansion.

Future Research Directions for Elasticity

While addressing the elasticity, the researchers should include the scenarios of busy workloads and should adopt different forecast methods. Then, only the improved performance and low-cost results can be achieved. We can try new scenarios, including the load balancing objectives, along with elasticity. We can stress adaptively using more virtual machines. Another way to increase the performance is by using the SLA protocols, along with elasticity using cost-effective approaches [112]. In future, researchers can plan to extend the elasticity with queuing theory-based model, “where the server is treated as a queuing framework and its theoretic results are used to derive a relationship between the request rate, service times of requests, and the response time SLA”, for the estimation of capacity regarding provisioning on the cloud [158]. Some of the latest work in elasticity include Refs. [159,160].

7.3. Consistency: Challenges and Issues

Maintaining consistency will enhance the replication strategies to a great extent. The primary importance should be always given to data integrity and consistency in a replicated domain for high performance. There is always a requirement for a strict consistency and strict consistency is the need of a high precision applications [34]. A consistency model in distributed domain figures out which guarantees can be expected for an update operation and for accessing an updated object. Its open challenges in cloud computing architecture to obtaining the correct balance between higher levels of consistency and availability [128]. However, more replication increases inconsistent replicas and strong replication (traditional synchronous) has its restrictions because of deficient performance and latency. Another important factor to hinder the strong replication in clouds is its geographically distant factor. Moreover, frequent data updates occur in clouds, which makes it burdensome to maintain the consistency of the replicas among the entire cloud [26].

Future Research Directions for Consistency

To achieve a strong consistency in cloud dispenses higher downtime because latencies become more prominent with strong consistency. Strong consistency is expensive not just in the transactional cost but also in terms of replicas availability and system performance [105]. Consequently, cloud storage systems have moved to eventual consistency

(all replicas eventually receive all writes). The major advantages of eventual consistency are performance, high availability and still provides a good enough consistency guarantee for production systems. However, to maintain both the availability and the performance following consistency is too costly. As the number of users increases (more users deliver more updates) on the cloud, there will be more stale data (probably two out of three reads are useless) which gradually decrease the performance. Therefore, there is a strong need to maintain high availability and consistency while not degrading the performance [26]. In this regard, the researchers should pay attention to maintain a balance between consistency, availability, and performance using adaptive methods [105,106] for better consistency. Some of the latest work in consistency include Refs. [161,162].

7.4. Cost: Challenges and Issues

While talking about the economic aspect of the replication strategies in cloud systems, cost plays a vital role as it acts as the most important objective while choosing any replication strategy. Desirable System performance must always be obtained at an acceptable cost [73]. There are various types of costs associated with replication strategies. Some of them are related to storage (data storage or data transfer costs), which rely on replica time, replica quantity, replica selection, and data movement [9,83,95], some are related to QoS and are based on mutual agreements [93], and some are related to monetary cost while considering tenant and providers benefits [115].

Future Research Directions for Cost

Cost is a very important attribute which needs further discussion due to its direct effect on sustainability and economic aspects of cloud systems. Cost and its utilization in high processing systems should be given a prime priority as all types of replication costs are directly associated with end-users and their service provider. Perhaps, the replication cost should increase provider and users benefits for performance guarantee. There is a strong demand to utilize lesser replication costs while not degrading the performance.

One of the future directions can be balancing an optimal number of tenants through the pay as you go, model, while satisfying the response time attribute resulting in an optimal profit for the provider [115]. The other future direction can be the implementation of these cost-based strategies of replications while taking into consideration of the energy consumption [116].

Reference [163] has mentioned some of the valuable future directions. Besides all enhancements, these cost-based replication strategies should be implemented in a real cloud environment [116,117]. Some of the latest work in cost include Refs. [164,165].

Figure 8 depicts the future research directions of least addressed target objective replication strategies in clouds.

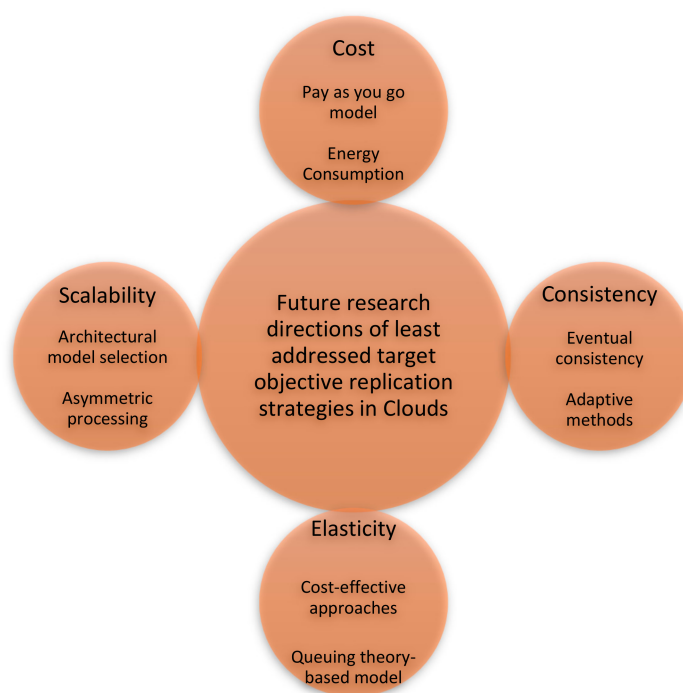


Figure 8. Future research directions of least addressed target objective replication strategies in clouds.

8. Discussion

As proposed in Section 4, the target based dynamic replication taxonomy in cloud configuration provides the depth in understanding the target objectives in the form of Primary target objective, Secondary target objective, and Tertiary target objectives based on most addressed, average addressed and least addressed objectives. Most of the targets of target-oriented replication strategies have concentrated on various target objectives, such as data availability, followed by reliability, and performance. These three target objectives are called primary target-based objectives as they are mostly addressed. The taxonomy also mentioned the secondary target objectives, such as fault tolerance and load balancing, as they are average addressed and additionally also mentioned the tertiary based objectives, such as scalability, elasticity, consistency, and cost, as they are least addressed. Table 4 represents the relationship of various dynamic replication strategies with their target objectives based on their attributes, purpose, and metrics. These attributes act as important metrics for the evaluation of target-oriented replication strategies in the cloud. Distinct target-oriented replication strategies cover different parameters based on their target objectives (either directly or indirectly). These attributes act as vital metrics for the evaluation of target-oriented replication strategies in the cloud. Table 5 represents the quantitative analysis of all target objectives in detail in the form of summary of all target objectives. Table 3 represents the literature review of target-oriented replication strategies.

In Section 5, a complete performance evaluation of different target objectives was performed in detail, along with feature comparison. We provide a comparative analysis and evaluation of various strategy in cloud computing environment, shown in Table 6 and Figure 6. Table 6 shows how various research papers have considered different parameters and discuss the impact of each strategies on target objectives. After reviewing the various target-oriented replication strategies comprehensively, it can be stated that different strategies have considered different metrics for evaluation. Concerning to target objectives, each strategy may consider one or multiple targets. Some of the strategies have considered a single target objective while some have included multiple target objectives for their metrics. Table 7 shows the feature of each respective target objective replication strategies with the intensities.

9. Conclusions

Replication strategies have been widely adopted in current cloud systems for data availability, reliability, and performance. The adaptation improves system resilience during disasters without any downtime. The cloud replication strategy trend to preserve the geographically distributed huge data, hence, creates the need for optimal replication strategy for acceptable performance. We filter out the dynamic replication strategies and evaluate their optimization capabilities based on quantitative analysis of target objectives (Primary target objective, Secondary target objective, and Tertiary target objective) using different attributes that are addressed. We provide a critical quantitative analysis and a comprehensive performance evaluation based on target objectives. We perform a comparative parameter evaluation, along with the metrics comparison. The paper also discusses the challenges, issues, and future research directions. This study will be beneficial to researchers to identify the research problems of replication strategies in cloud computing configuration and will provide a depth in detail related to available dynamic replication strategies and target-oriented replication strategies. This research will open a new gate to develop the optimal dynamic replication strategy for clouds in the future.

Author Contributions: The authors of this article have contributed to this research paper as follows: Writing and preparation, Q.W.; Review and visualization, W.I.S.W.D., S.S.A., A.A., and A.N.; Editing and revision, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: Taif University Researchers Supporting Project number (TURSP-2020/215), Taif University, Taif, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hameed, A.; Khoshkbarforousha, A.; Ranjan, R.; Jayaraman, P.P.; Kolodziej, J.; Balaji, P.; Zeadally, S.; Malluhi, Q.M.; Tziritas, N.; Vishnu, A.; et al. A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. *Computing* **2014**, *98*, 751–774. [\[CrossRef\]](#)
2. Yavari, M.; Rahbar, A.G.; Fathi, M.H. Temperature and energy-aware consolidation algorithms in cloud computing. *J. Cloud Comput.* **2019**, *8*, 1–16. [\[CrossRef\]](#)
3. Mansouri, N.; Rafsanjani, M.K.; Javidi, M. DPRS: A dynamic popularity aware replication strategy with parallel download scheme in cloud environments. *Simul. Model. Pract. Theory* **2017**, *77*, 177–196. [\[CrossRef\]](#)
4. Ebadi, Y.; Navimipour, N.J. An energy-aware method for data replication in the cloud environments using a Tabu search and particle swarm optimization algorithm. *Concurr. Comput. Pract. Exp.* **2019**, *31*, e4757. [\[CrossRef\]](#)
5. Milani, B.A.; Nima, J.N. A comprehensive review of the data replication techniques in the cloud environments: Major trends and future directions. *J. Netw. Comput. Appl.* **2016**, *64*, 229–238. [\[CrossRef\]](#)
6. Zhao, L.; Sakr, S.; Liu, A.; Bouguettaya, A. SLA-Driven Database Replication on Virtualized Database Servers. In *Cloud Data Management*; Springer: Cham, Switzerland, 2014; pp. 97–118.
7. Malik, S.U.R.; Khan, S.U.; Ewen, S.J.; Tziritas, N.; Kolodziej, J.; Zomaya, A.Y.; Li, H. Performance analysis of data intensive cloud systems based on data management and replication: A survey. *Distrib. Parallel Databases* **2016**, *34*, 179–215. [\[CrossRef\]](#)
8. Ikeda, T.; Ohara, M.; Fukumoto, S.; Arai, M.; Iwasaki, K. A Distributed Data Replication Protocol for File Versioning with Optimal Node Assignments. In Proceedings of the 2010 IEEE 16th Pacific Rim International Symposium on Dependable Computing, Tokyo, Japan, 13–15 December 2010; pp. 117–124.
9. Lin, J.-W.; Chen, C.-H.; Chang, J.M. QoS-aware data replication for data-intensive applications in cloud computing systems. *IEEE Trans. Cloud Comput.* **2013**, *1*, 101–115.
10. Fazilina, A.; Latip, R.; Ibrahim, H.; Abdullah, A. A Review: Replication Strategies for Big Data in Cloud Environment. *Int. J. Eng. Technol.* **2018**, *7*, 357–362.
11. Tomar, D.; Tomar, P. Integration of Cloud Computing and Big Data Technology for Smart Generation. In Proceedings of the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 11–12 January 2018; pp. 1–6.
12. Xia, Q.; Liang, W.; Xu, Z. QoS-Aware data replications and placements for query evaluation of big data analytics. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017; pp. 1–7.
13. Gopinath, S.; Sherly, E. A Weighted Dynamic Data Replication Management for Cloud Data Storage Systems. *Int. J. Appl. Eng. Res.* **2017**, *12*, 15517–15524.
14. Li, Y.; Yu, M.; Xu, M.; Yang, J.; Sha, D.; Liu, Q.; Yang, C. Big Data and Cloud Computing. In *Manual of Digital Earth*; Guo, H., Goodchild, M.F., Annoni, A., Eds.; Springer: Singapore, 2020.

15. Das, M.; Dash, R. Role of Cloud Computing for Big Data: A Review. In *Intelligent and Cloud Computing. Smart Innovation, Systems and Technologies*; Mishra, D., Buyya, R., Mohapatra, P., Patnaik, S., Eds.; Springer: Singapore, 2021.
16. Khan, S.; Shakil, K.A.; Alam, M.; Aggarwal, V.B.; Bhatnagar, V.; Mishra, D.K. Cloud-Based Big Data Analytics—A Survey of Current Research and Future Directions. *Adv. Intell. Syst. Comput.* **2017**, 595–604. [\[CrossRef\]](#)
17. Kobusińska, A.; Leung, C.; Hsu, C.-H.; Raghavendra, S.; Chang, V. Emerging trends, issues and challenges in Internet of Things, Big Data and cloud computing. *Futur. Gener. Comput. Syst.* **2018**, 87, 416–419. [\[CrossRef\]](#)
18. Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big Data and cloud computing: Innovation opportunities and challenges. *Int. J. Digit. Earth* **2017**, 10, 13–53. [\[CrossRef\]](#)
19. Rao, T.R.; Mitra, P.; Bhatt, R.; Goswami, A. The big data system, components, tools, and technologies: A survey. *Knowl. Inf. Syst.* **2019**, 60, 1165–1245. [\[CrossRef\]](#)
20. Nachiappan, R.; Javadi, B.; Calheiros, R.N.; Matawie, K.M. Cloud storage reliability for Big Data applications: A state of the art survey. *J. Netw. Comput. Appl.* **2017**, 97, 35–47. [\[CrossRef\]](#)
21. Hashem IA, T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Khan, S.U. The rise of “big data” on cloud computing: Review and open research issues. *Inf. Syst.* **2015**, 47, 98–115. [\[CrossRef\]](#)
22. Aceto, G.; Persico, V.; Pescapé, A. Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for Healthcare 4.0. *J. Ind. Inf. Integr.* **2020**, 18, 100129. [\[CrossRef\]](#)
23. Tahir, A.; Chen, F.; Khan, H.U.; Ming, Z.; Ahmad, A.; Nazir, S.; Shafiq, M. A Systematic Review on Cloud Storage Mechanisms Concerning e-Healthcare Systems. *Sensors* **2020**, 20, 5392. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Shorfuzzaman, M.; Masud, M. Leveraging A Multi-Objective Approach to Data Replication in Cloud Computing Environment to Support Big Data Applications. *Int. J. Adv. Comput. Sci. Appl.* **2019**, 10. [\[CrossRef\]](#)
25. Gopinath, S.; Sherly, E. A Comprehensive Survey on Data Replication Techniques in Cloud Storage Systems. *Int. J. Appl. Eng. Res.* **2018**, 13, 15926–15932.
26. Chihoub, H.-E.; Ibrahim, S.; Antoniu, G.; Pérez, M.S. Harmony: Towards Automated Self-Adaptive Consistency in Cloud Storage. In Proceedings of the 2012 IEEE International Conference on Cluster Computing, Beijing, China, 24–28 September 2012; pp. 293–301.
27. Azimi, k.S. A Bee Colony (Beehive) based approach for data replication in cloud environments. In *Fundamental Research in Electrical Engineering: The Selected Papers of The First International Conference on Fundamental Research in Electrical Engineering*; Springer: Singapore, 2019; pp. 1039–1052.
28. Boru, D.; Kliazovich, D.; Granelli, F.; Bouvry, P.; Zomaya, A.Y. Energy-efficient data replication in cloud computing datacenters. *Clust. Comput.* **2015**, 18, 385–402. [\[CrossRef\]](#)
29. Abadi, D.J. Data management in the cloud: Limitations and opportunities. *IEEE Data Eng. Bull.* **2009**, 32, 3–12.
30. Amjad, T.; Sher, M.; Daud, A. A survey of dynamic replication strategies for improving data availability in data grids. *Futur. Gener. Comput. Syst.* **2012**, 28, 337–349. [\[CrossRef\]](#)
31. Karandikar, R.; Manish, G. Analytical Survey of Dynamic Replication Strategies in Cloud. In Proceedings of the IJCA-National Conference on Recent Trends in Computer Science and Information Technology, Nagpur, India, 1–5 June 2016.
32. Hamrouni, T.; Sarra, S.; Charrada, F.B. A survey of dynamic replication and replica selection strategies based on data mining techniques in data grids. *Eng. Appl. Artif. Intell.* **2016**, 48, 140–158. [\[CrossRef\]](#)
33. Pan, S.; Xiong, L.; Xu, Z.; Chong, Y.; Meng, Q. A dynamic replication management strategy in distributed GIS. *Comput. Geosci.* **2018**, 112, 1–8. [\[CrossRef\]](#)
34. Goel, S.; Rajkumar, B. Data replication strategies in wide-area distributed systems. In *Enterprise Service Computing: From Concept to Deployment*; IGI Global: Hershey, PA, USA, 2007; pp. 211–241.
35. Milani, B.A.; Navimipour, N.J. A Systematic Literature Review of the Data Replication Techniques in the Cloud Environments. *Big Data Res.* **2017**, 10, 1–7. [\[CrossRef\]](#)
36. Warhade, S.; Dahiwal, P.; Raghuvanshi, M. A Dynamic Data Replication in Grid System. *Procedia Comput. Sci.* **2016**, 78, 537–543. [\[CrossRef\]](#)
37. Naseera, S. A survey on data replication strategies in a Data Grid environment. *Multiagent Grid Syst.* **2017**, 12, 253–269. [\[CrossRef\]](#)
38. Vashisht, P.; Anju, S.; Rajesh, K. Strategies for replica consistency in data grid—A comprehensive survey. *Concurr. Comput. Pract. Exp.* **2017**, 29, e3907. [\[CrossRef\]](#)
39. Tos, U.; Mokadem, R.; Hameurlain, A.; Ayav, T.; Bora, S. Dynamic replication strategies in data grid systems: A survey. *J. Supercomput.* **2015**, 71, 4116–4140. [\[CrossRef\]](#)
40. Hamrouni, T.; Slimani, S.; Ben Charrada, F. A Critical Survey of Data Grid Replication Strategies Based on Data Mining Techniques. *Procedia Comput. Sci.* **2015**, 51, 2779–2788. [\[CrossRef\]](#)
41. Mansouri, N.; Javidi, M.M. A Survey of Dynamic Replication Strategies for Improving Response Time in Data Grid Environment. *Amirkabir Int. J. ModelingIdentif. Simul. Control* **2017**, 49, 239–264.
42. Souravlas, S.; Sifaleras, A. Trends in data replication strategies: A survey. *Int. J. Parallel Emergent Distrib. Syst.* **2019**, 34, 222–239. [\[CrossRef\]](#)
43. Vashisht, P.; Kumar, V.; Kumar, R.; Sharma, A. Optimizing Replica Creation using Agents in Data Grids. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 4–6 February 2019; pp. 542–547.

44. Hamrouni, T.; Hamdeni, C.; Ben Charrada, F. Objective assessment of the performance of data grid replication strategies based on distribution quality. *Int. J. Web Eng. Technol.* **2016**, *11*, 3–28. [\[CrossRef\]](#)
45. Hamrouni, T. Replication in Data Grids: Metrics and Strategies. *arXiv* **2019**, arXiv:1912.10171.
46. Lwin, T.K.; Alexander, B. Real time analysis of data grid processing for future technology. In Proceedings of the International Conference on Computer Science and Information Technologies, Yerevan, Armenia, 23–27 September 2019; pp. 53–54.
47. Salah, T.; Zemerly, M.J.; Yeun, C.Y.; Al-Qutayri, M.; Al-Hammadi, Y. The evolution of distributed systems towards microservices architecture. In Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 22 April 2016; pp. 318–325.
48. Mokadem, R.; Hameurlain, A. Data replication strategies with performance objective in data grid systems: A survey. *Int. J. Grid Util. Comput.* **2015**, *6*, 30. [\[CrossRef\]](#)
49. Yang, C.; Clarke, K.; Shekhar, S.; Tao, C.V. Big Spatiotemporal Data Analytics: A research and innovation frontier. *Int. J. Geogr. Inf. Sci.* **2019**, *34*, 1075–1088. [\[CrossRef\]](#)
50. Spaho, E.; Barolli, L.; Khafa, F. Data Replication Strategies in P2P Systems: A Survey. In Proceedings of the 2014 17th International Conference on Network-Based Information Systems, Salerno, Italy, 10–12 September 2014; pp. 302–309.
51. Sun, S.; Yao, W.; Qiao, B.; Zong, M.; He, X.; Li, X. RRSD: A file replication method for ensuring data reliability and reducing storage consumption in a dynamic Cloud-P2P environment. *Futur. Gener. Comput. Syst.* **2019**, *100*, 844–858. [\[CrossRef\]](#)
52. Tabet, K.; Mokadem, R.; Laouar, M.R. A data replication strategy for document-oriented NoSQL systems. *Int. J. Grid Util. Comput.* **2019**, *10*, 53–62. [\[CrossRef\]](#)
53. Wang, S.; Batiha, K. A metaheuristic-based method for replica selection in the Internet of Things. *Int. J. Commun. Syst.* **2020**, *33*, e4458. [\[CrossRef\]](#)
54. Lazeb, A.; Mokadem, R.; Belalem, G. Towards a New Data Replication Management in Cloud Systems. *Int. J. Strat. Inf. Technol. Appl.* **2019**, *10*, 1–20. [\[CrossRef\]](#)
55. Abdollahi, N.A.; Rajabion, L. Data replication techniques in the mobile ad hoc networks: A systematic and comprehensive review. *Int. J. Pervasive Comput. Commun.* **2019**, *15*, 174–198. [\[CrossRef\]](#)
56. Nassif, A.B.; Abu Talib, M.; Nasir, Q.; Albadani, H.; Dakalbab, F.M. Machine Learning for Cloud Security: A Systematic Review. *IEEE Access* **2021**, *9*, 20717–20735. [\[CrossRef\]](#)
57. Pan, Q.; Wu, J.; Zheng, X.; Li, J.; Li, S.; Vasilakos, A.V. Leveraging AI and Intelligent Reflecting Surface for Energy-Efficient Communication in 6G IoT. *arXiv* **2020**, arXiv:2012.14716.
58. Hasenburg, J.; Grambow, M.; Bermbach, D. Towards a replication service for data-intensive fog applications. In Proceedings of the 35th Annual ACM Symposium on Applied Computing; Association for Computing Machinery (ACM), Brno, Czech Republic, 30 March–3 April 2019; pp. 267–270.
59. Liu, X.; Xie, L.; Wang, Y.; Zou, J.; Xiong, J.; Ying, Z.; Vasilakos, A.V. Privacy and Security Issues in Deep Learning: A Survey. *IEEE Access* **2021**, *9*, 4566–4593. [\[CrossRef\]](#)
60. Hamdan, M.; Hassan, E.; Abdelaziz, A.; Elhigazi, A.; Mohammed, B.; Khan, S.; Vasilakos, A.V.; Marsono, M. A comprehensive survey of load balancing techniques in software-defined network. *J. Netw. Comput. Appl.* **2021**, *174*, 102856. [\[CrossRef\]](#)
61. Ni, J.; Zhang, K.; Vasilakos, A.V. Security and Privacy for Mobile Edge Caching: Challenges and Solutions. *IEEE Wirel. Commun.* **2020**, 1–7. [\[CrossRef\]](#)
62. Mansouri, N.; Javidi, M.M.; Zade, B.M.H. A CSO-based approach for secure data replication in cloud computing environment. *J. Supercomput.* **2020**, 1–52. [\[CrossRef\]](#)
63. Alam, M.; Mazliham, M.; Yeakub, M. A Survey of Machine Learning Algorithms in Cloud Computing. In *The Perspective of Network Data Replication Decision*; UniKL Postgraduate Symposium: Kuala Lumpur, Malaysia, 2013.
64. Kale, R.V.; Veeravalli, B.; Wang, X. A Practicable Machine Learning Solution for Security-Cognizant Data Placement on Cloud Platforms. In *Handbook of Computer Networks and Cyber Security*; Springer: Cham, Switzerland, 2020; pp. 111–131.
65. Tabet, K.; Mokadem, R.; Laouar, M.R.; Eom, S. Data replication in cloud systems: A survey. *Int. J. Inf. Syst. Soc. Chang.* **2017**, *8*, 17–33. [\[CrossRef\]](#)
66. Bhuvaneshwari, R.; Ravi, T. A Review of Static and Dynamic Data Replication Mechanisms for Distributed Systems. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 953–964. [\[CrossRef\]](#)
67. Edwin, E.B.; Umamaheswari, P.; Thanka, M.R. An efficient and improved multi-objective optimized replication management with dynamic and cost aware strategies in cloud computing data center. *Clust. Comput.* **2017**, *22*, 11119–11128. [\[CrossRef\]](#)
68. Mealha, D.; Preguiça, N.; Gomes, M.C.; Leitão, J. Data Replication on the Cloud/Edge. In Proceedings of the 6th Workshop on Principles and Practice of Consistency for Distributed Data—PaPoC'19, Dresden Germany, 28–25 March 2019; pp. 1–7.
69. Saranya, N.; Geetha, K.; Rajan, C. Data Replication in Mobile Edge Computing Systems to Reduce Latency in Internet of Things. *Wirel. Pers. Commun.* **2020**, *112*, 2643–2662. [\[CrossRef\]](#)
70. Atrey, A.; Van Seghbroeck, G.; Mora, H.; De Turck, F.; Volckaert, B. Unifying Data and Replica Placement for Data-intensive Services in Geographically Distributed Clouds. In Proceedings of the Proceedings of the 9th International Conference on Cloud Computing and Services Science, Heraklion, Greece, 2–4 May 2019; pp. 25–36.
71. Lazeb, A.; Mokadem, R.; Belalem, G. Economic Data Replication Management in the Cloud. In *JERI*; Saida, Algeria, 27 April 2019. Available online: <https://www.semanticscholar.org/paper/Economic-Data-Replication-Management-in-the-Cloud-Lazeb-Mokadem/49a10747912ff82f69d5685ed4181751e92aa9a8> (accessed on 9 March 2021).

72. Slimani, S.; Hamrouni, T.; Ben Charrada, F.; Magoules, F. DDSor: A Dependency Aware Dynamic Service Replication Strategy for Efficient Execution of Service-Oriented Applications in the Cloud. In Proceedings of the 2017 International Conference on High Performance Computing & Simulation (HPCS), Genoa, Italy, 17–21 July 2017; pp. 603–610.
73. Slimani, S.; Hamrouni, T.; Ben Charrada, F. Service-oriented replication strategies for improving quality-of-service in cloud computing: A survey. *Clust. Comput.* **2021**, *24*, 361–392. [\[CrossRef\]](#)
74. Mohamed, M.F. Service replication taxonomy in distributed environments. *Serv. Oriented Comput. Appl.* **2016**, *10*, 317–336. [\[CrossRef\]](#)
75. Björkqvist, M.F.; Chen, L.Y.; Binder, W. Dynamic Replication in Service-Oriented Systems. In Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (Ccgird 2012), Ottawa, ON, Canada, 13–16 May 2012; pp. 531–538.
76. Wu, J.; Zhang, B.; Yang, L.; Wang, P.; Zhang, C. A replicas placement approach of component services for service-based cloud application. *Clust. Comput.* **2016**, *19*, 709–721. [\[CrossRef\]](#)
77. Chen, T.; Bahsoon, R.; Tawil, A.-R.H. Scalable service-oriented replication with flexible consistency guarantee in the cloud. *Inf. Sci.* **2014**, *264*, 349–370. [\[CrossRef\]](#)
78. Tos, U.; Mokadem, R.; Hameurlain, A.; Ayav, T.; Bora, S. A Performance and Profit Oriented Data Replication Strategy for Cloud Systems. In Proceedings of the 2016 International IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld), Toulouse, France, 18–21 July 2016; pp. 780–787.
79. Mesbahi, M.; Rahmani, A.M. Load Balancing in Cloud Computing: A State of the Art Survey. *Int. J. Mod. Educ. Comput. Sci.* **2016**, *8*, 64–78. [\[CrossRef\]](#)
80. You, X.; Li, Y.; Zheng, M.; Zhu, C.; Yu, L. A Survey and Taxonomy of Energy Efficiency Relevant Surveys in Cloud-Related Environments. *IEEE Access* **2017**, *5*, 14066–14078. [\[CrossRef\]](#)
81. Ali, S.A.; Affan, M.; Alam, M. A study of efficient energy management techniques for cloud computing environment. *arXiv* **2018**, arXiv:1810.07458.
82. Huang, H.; Hung, W.; Shin, K.G. FS2: Dynamic data replication in free disk space for improving disk performance and energy consumption. *Acm Sigops Oper. Syst. Rev.* **2005**, *39*, 263–276. [\[CrossRef\]](#)
83. Singh, L.; Malhotra, J. A Survey on Data Placement Strategies for Cloud based Scientific Workflows. *Int. J. Comput. Appl.* **2016**, *141*, 30–33. [\[CrossRef\]](#)
84. Wei, Q.; Veeravalli, B.; Gong, B.; Zeng, L.; Feng, D. CDRM: A Cost-Effective Dynamic Replication Management Scheme for Cloud Storage Cluster. In Proceedings of the 2010 IEEE International Conference on Cluster Computing; Institute of Electrical and Electronics Engineers (IEEE), Heraklion, Greece, 20–24 September 2010; pp. 188–196.
85. Mansouri, N.; Javidi, M.M. A review of data replication based on meta-heuristics approach in cloud computing and data grid. *Soft Comput.* **2020**, *24*, 1–28. [\[CrossRef\]](#)
86. Cheng, Z.; Luan, Z.; Meng, Y.; Xu, Y.; Qian, D.; Roy, A.; Zhang, N.; Guan, G. ERMS: An Elastic Replication Management System for HDFS. In Proceedings of the 2012 IEEE International Conference on Cluster Computing Workshops, Beijing, China, 24–28 September 2012; pp. 32–40.
87. Bui, D.-M.; Hussain, S.; Huh, E.-N.; Lee, S. Adaptive Replication Management in HDFS Based on Supervised Learning. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 1369–1382. [\[CrossRef\]](#)
88. Qu, K.; Meng, L.; Yang, Y. A dynamic replica strategy based on Markov model for hadoop distributed file system (HDFS). In Proceedings of the 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing, China, 17–19 August 2016; pp. 337–342.
89. Kousiouris, G.; Vafiadis, G.; Varvarigou, T. Enabling Proactive Data Management in Virtualized Hadoop Clusters Based on Predicted Data Activity Patterns. In Proceedings of the 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Compiègne, France, 28–30 October 2013; pp. 1–8.
90. Rimal, B.P.; Jukan, A.; Katsaros, D.; Goeleven, Y. Architectural Requirements for Cloud Computing Systems: An Enterprise Cloud Approach. *J. Grid Comput.* **2011**, *9*, 3–26. [\[CrossRef\]](#)
91. Zhang, T. A QoS-enhanced data replication service in virtualised cloud environments. *Int. J. Netw. Virtual Organ.* **2020**, *22*, 1–16. [\[CrossRef\]](#)
92. Faraidoon, H.; Nagesh, K. A brief survey on dynamic strategies of data replication in cloud environment: Last five year study. *Int. J. Eng. Dev. Res.* **2017**, *5*, 342–345.
93. Zia, A.; Khan, M.N.A. Identifying Key Challenges in Performance Issues in Cloud Computing. *Int. J. Mod. Educ. Comput. Sci.* **2012**, *4*, 59–68. [\[CrossRef\]](#)
94. Xia, Q.; Bai, L.; Liang, W.; Xu, Z.; Yao, L.; Wang, L. Qos-aware proactive data replication for big data analytics in edge clouds. In Proceedings of the 48th International Conference on Parallel Processing: Workshops, Kyoto, Japan, 5–8 August 2019; pp. 1–10.
95. Xia, Q.; Xu, Z.; Liang, W.; Yu, S.; Guo, S.; Zomaya, A.Y. Efficient Data Placement and Replication for QoS-Aware Approximate Query Evaluation of Big Data Analytics. *IEEE Trans. Parallel Distrib. Syst.* **2019**, *30*, 2677–2691. [\[CrossRef\]](#)
96. Kumar, P.J.; Ilango, P. BMAQR: Balanced multi attribute QoS aware replication in HDFS. *Int. J. Internet Technol. Secur. Trans.* **2018**, *8*, 195–208. [\[CrossRef\]](#)

97. Chauhan, N.; Tripathi, S.P. QoS Aware Replica Control Strategies for Distributed Real Time Database Management System. *Wirel. Pers. Commun.* **2018**, *104*, 739–752. [\[CrossRef\]](#)
98. Long, S.-Q.; Zhao, Y.-L.; Chen, W. MORM: A Multi-objective Optimized Replication Management strategy for cloud storage cluster. *J. Syst. Arch.* **2014**, *60*, 234–244. [\[CrossRef\]](#)
99. Xie, F.; Yan, J.; Shen, J. Towards Cost Reduction in Cloud-Based Workflow Management through Data Replication. In Proceedings of the 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), Shanghai, China, 13–14 August 2017; pp. 94–99.
100. Li, W.; Yang, Y.; Yuan, D. A Novel Cost-Effective Dynamic Data Replication Strategy for Reliability in Cloud Data Centres. In Proceedings of the 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure, Sydney, NSW, Australia, 12–14 December 2011; pp. 496–502.
101. Sun, D.-W.; Chang, G.-R.; Gao, S.; Jin, L.-Z.; Wang, X.-W. Modeling a Dynamic Data Replication Strategy to Increase System Availability in Cloud Computing Environments. *J. Comput. Sci. Technol.* **2012**, *27*, 256–272. [\[CrossRef\]](#)
102. Li, W.; Yang, Y.; Yuan, D. Ensuring Cloud Data Reliability with Minimum Replication by Proactive Replica Checking. *IEEE Trans. Comput.* **2015**, *65*, 1494–1506. [\[CrossRef\]](#)
103. Chihoub, H.-E.; Ibrahim, S.; Antoniu, G.; Perez, M.S. Consistency in the Cloud: When Money Does Matter! In Proceedings of the 2013 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, Delft, The Netherlands, 13–16 May 2013; pp. 352–359.
104. Hussein, M.-K.; Mousa, M. A light-weight data replication for cloud data centers environment. *Int. J. Eng. Innov. Technol.* **2012**, *1*, 169–175.
105. Radi, M. Runtime Replica Consistency Mechanism for Cloud Data Storage. In Proceedings of the International Conference on Information & Communication Technology: Application & Techniques (ICICT 2012), Ramallah, Palestine, 16–17 April 2012.
106. Phansalkar, S.P.; Dani, A.R. Tunable consistency guarantees of selective data consistency model. *J. Cloud Comput.* **2015**, *4*, 13. [\[CrossRef\]](#)
107. Gill, N.K.; Singh, S. A dynamic, cost-aware, optimized data replication strategy for heterogeneous cloud data centers. *Futur. Gener. Comput. Syst.* **2016**, *65*, 10–32. [\[CrossRef\]](#)
108. Bai, X.; Jin, H.; Liao, X.; Shi, X.; Shao, Z. RTRM: A Response Time-Based Replica Management Strategy for Cloud Storage System. In Proceedings of the Constructive Side-Channel Analysis and Secure Design, Paris, France, 6–8 March 2013; pp. 124–133.
109. Kirubakaran, S.; Valarmathy, S.; Kamalanathan, C. Data replication using modified D2RS in cloud computing for performance improvement. *J. Theor. Appl. Inf. Technol.* **2013**, *pp*, 460–470.
110. Rajalakshmi, A.; Vijayakumar, D.; Srinivasagan, K.G. An improved dynamic data replica selection and placement in cloud. In Proceedings of the 2014 International Conference on Recent Trends in Information Technology, Chennai, India, 10–12 April 2014; pp. 1–6.
111. Xue, M.; Jing, S.J.; Feng, G.X. Replica Placement in Cloud Storage based on Minimal Blocking Probability. In Proceedings of the The 5th International Conference on Computer Engineering and Networks, Shanghai, China, 12–13 September 2015.
112. Sousa, F.R.C.; Moreira, L.O.; Filho, J.S.C.; Machado, J.C. Predictive elastic replication for multi-tenant databases in the cloud. *Concurr. Comput. Pr. Exp.* **2018**, *30*, e4437. [\[CrossRef\]](#)
113. Mansouri, N. Adaptive data replication strategy in cloud computing for performance improvement. *Front. Comput. Sci.* **2016**, *10*, 925–935. [\[CrossRef\]](#)
114. Sun, S.; Yao, W.; Li, X. DARS: A dynamic adaptive replica strategy under high load Cloud-P2P. *Futur. Gener. Comput. Syst.* **2018**, *78*, 31–40. [\[CrossRef\]](#)
115. Mokadem, R.; Hameurlain, A. A data replication strategy with tenant performance and provider economic profit guarantees in Cloud data centers. *J. Syst. Softw.* **2020**, *159*, 110447. [\[CrossRef\]](#)
116. Limam, S.; Mokadem, R.; Belalem, G. Data replication strategy with satisfaction of availability, performance and tenant budget requirements. *Clust. Comput.* **2019**, *22*, 1199–1210. [\[CrossRef\]](#)
117. Tos, U.; Mokadem, R.; Hameurlain, A.; Ayav, T.; Bora, S. Ensuring performance and provider profit through data replication in cloud systems. *Clust. Comput.* **2017**, *21*, 1479–1492. [\[CrossRef\]](#)
118. Tu, M.; Xiao, L.; Xu, D. Maximizing the Availability of Replicated Services in Widely Distributed Systems Considering Network Availability. In Proceedings of the 2013 IEEE 7th International Conference on Software Security and Reliability, Gaithersburg, MD, USA, 18–20 June 2013; pp. 178–187.
119. Bachwani, R.; Gryz, L.; Bianchini, R.; Dubnicki, C. Dynamically Quantifying and Improving the Reliability of Distributed Storage Systems. In Proceedings of the 2008 Symposium on Reliable Distributed Systems, Naples, Italy, 6–8 October 2008; pp. 85–94.
120. Yang, J.-P. Efficient Load Balancing Using Active Replica Management in a Storage System. *Math. Probl. Eng.* **2016**, *2016*, 1–9. [\[CrossRef\]](#)
121. Oo, M.; Soe, T.T.; Thida, A. Fault tolerance by replication of distributed database in P2P system using agent approach. *Int. J. Comput.* **2010**, *4*, 9–18.
122. Amoon, M. Adaptive Framework for Reliable Cloud Computing Environment. *IEEE Access* **2016**, *4*, 9469–9478. [\[CrossRef\]](#)
123. Amiri, M.J.; Maiyya, S.; Agrawal, D.; El Abbadi, A. SeeMoRe: A Fault-Tolerant Protocol for Hybrid Cloud Environments. In Proceedings of the 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, 20–24 April 2020; pp. 1345–1356.

124. Kumar, P.; Kumar, R. Issues and challenges of load balancing techniques in cloud computing: A survey. *Acm Comput. Surv.* **2019**, *51*, 1–35. [\[CrossRef\]](#)
125. Thakur, A.; Goraya, M.S. A taxonomic survey on load balancing in cloud. *J. Netw. Comput. Appl.* **2017**, *98*, 43–57. [\[CrossRef\]](#)
126. Lehrig, S.; Sanders, R.; Brataas, G.; Cecowski, M.; Ivanšek, S.; Polutnik, J. CloudStore—Towards scalability, elasticity, and efficiency benchmarking and analysis in Cloud computing. *Futur. Gener. Comput. Syst.* **2018**, *78*, 115–126. [\[CrossRef\]](#)
127. Patibandla, R.S.M.L.; Kurra, S.S.; Mundukur, N.B. A Study on Scalability of Services and Privacy Issues in Cloud Computing. Distributed Computing and Internet Technology. ICDIT 2012. Lecture Notes in Computer Science; Ramanujam, R., Ramaswamy, S., Eds.; Springer: Berlin, Heidelberg, Germany, 2012; 7154, pp. 212–230. [\[CrossRef\]](#)
128. Campêlo, R.A.; Casanova, M.A.; Guedes, D.O.; Laender, A.H.F. A brief survey on replica consistency in cloud environments. *J. Internet Serv. Appl.* **2020**, *11*, 1–13. [\[CrossRef\]](#)
129. Hassan, O.A.-H.; Ramaswamy, L.; Miller, J.; Rasheed, K.; Canfield, E.R. Replication in Overlay Networks: A Multi-objective Optimization Approach. In Proceedings of the International Conference on Collaborative Computing: Networking, Applications and Worksharing; Springer: Berlin, Heidelberg, Germany, 2008; pp. 512–528.
130. Perez-Sorrosal, F.; Patiño-Martínez, M.; Jimenez-Peris, R.; Kemme, B. Elastic SI-Cache: Consistent and scalable caching in multi-tier architectures. *VLDB J.* **2011**, *20*, 841–865. [\[CrossRef\]](#)
131. Qu, C.; Calheiros, R.N.; Buyya, R. Auto-scaling web applications in clouds: A taxonomy and survey. *ACM Comput. Surv.* **2018**, *51*, 1–33. [\[CrossRef\]](#)
132. Khelaifa, A.; Benharzallah, S.; Kahloul, L.; Euler, R.; Laouid, A.; Bounceur, A. A comparative analysis of adaptive consistency approaches in cloud storage. *J. Parallel Distrib. Comput.* **2019**, *129*, 36–49. [\[CrossRef\]](#)
133. Wu, W.; Wang, W.; Fang, X.; Junzhou, L.; Vasilakos, A.V. Electricity Price-aware Consolidation Algorithms for Time-sensitive VM Services in Cloud Systems. *IEEE Trans. Serv. Comput.* **2019**. [\[CrossRef\]](#)
134. Zeng, L.; Xu, S.; Wang, Y.; Kent, K.B.; Bremner, D.; Xu, C. Toward cost-effective replica placements in cloud storage systems with QoS-awareness. *Softw. Pract. Exp.* **2017**, *47*, 813–829. [\[CrossRef\]](#)
135. Casas, I.; Taheri, J.; Ranjan, R.; Wang, L.; Zomaya, A.Y. A balanced scheduler with data reuse and replication for scientific workflows in cloud computing systems. *Futur. Gener. Comput. Syst.* **2017**, *74*, 168–178. [\[CrossRef\]](#)
136. Statovci-Halimi, B.; Halimi, A. QoS management through service level agreements: A short overview. *Elektrotechnik Inf.* **2004**, *121*, 243–246. [\[CrossRef\]](#)
137. Ardagna, D.; Casale, G.; Ciavotta, M.; Pérez, J.F.; Wang, W. Quality-of-service in cloud computing: Modeling techniques and their applications. *J. Internet Serv. Appl.* **2014**, *5*, 11. [\[CrossRef\]](#)
138. Abad, C.L.; Lu, Y.; Campbell, R.H. DARE: Adaptive Data Replication for Efficient Cluster Scheduling. In Proceedings of the 2011 IEEE International Conference on Cluster Computing, Austin, TX, USA, 26–30 September 2011; pp. 159–168.
139. Chang, W.-C.; Wang, P.-C. Write-Aware Replica Placement for Cloud Computing. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 656–667. [\[CrossRef\]](#)
140. Vobugari, S.; Somayajulu, D.V.L.N.; Subaraya, B.M. Dynamic Replication Algorithm for Data Replication to Improve System Availability: A Performance Engineering Approach. *IETE J. Res.* **2015**, *61*, 132–141. [\[CrossRef\]](#)
141. Wei, J.; Liu, J.; Zhang, R.; Niu, X. Efficient Dynamic Replicated Data Possession Checking in Distributed Cloud Storage Systems. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 1894713. [\[CrossRef\]](#)
142. Guo, W.; Qin, S.; Lu, J.; Gao, F.; Jin, Z.; Wen, Q. Improved Proofs of Retrievability And Replication For Data Availability In Cloud Storage. *Comput. J.* **2020**, *63*, 1216–1230. [\[CrossRef\]](#)
143. Tos, U.; Mokadem, R.; Hameurlain, A.; Ayav, T. Achieving query performance in the cloud via a cost-effective data replication strategy. *Soft Comput.* **2021**, 1–18. [\[CrossRef\]](#)
144. John, S.N.; Mirnalinee, T.T. A novel dynamic data replication strategy to improve access efficiency of cloud storage. *Inf. Syst. E-Bus. Manag.* **2020**, *18*, 405–426. [\[CrossRef\]](#)
145. Karandikar, R.R.; Gudadhe, M. B Comparative analysis of dynamic replication strategies in cloud. *Int. J. Comput. Appl.* **2016**, *975*, 8887.
146. Miloudi, I.E.; Yagoubi, B.; Bellounar, F.Z. Dynamic Replication Based on a Data Classification Model in Cloud Computing. In *International Symposium on Modelling and Implementation of Complex Systems*; Springer: Cham, Switzerland, 24 October 2020; pp. 3–17.
147. Abbes, H.; Louati, T.; Cérin, C. Dynamic replication factor model for Linux containers-based cloud systems. *J. Supercomput.* **2020**, *76*, 7219–7241. [\[CrossRef\]](#)
148. Karuppusamy, S.; Muthaiyan, M. An Efficient Placement Algorithm for Data Replication and To Improve System Availability in Cloud Environment. *Int. J. Intell. Eng. Syst.* **2016**, *9*, 88–97. [\[CrossRef\]](#)
149. Wang, Y.; Wang, J. An Optimized Replica Distribution Method in Cloud Storage System. *J. Control. Sci. Eng.* **2017**, *2017*, 1–8. [\[CrossRef\]](#)
150. Mansouri, N.; Javidi, M.M.; Zade, B.M.H. Using data mining techniques to improve replica management in cloud environment. *Soft Comput.* **2020**, *24*, 7335–7360. [\[CrossRef\]](#)
151. Kaur, A.; Gupta, P.; Singh, M.; Nayyar, A. Data Placement in Era of Cloud Computing: A Survey, Taxonomy and Open Research Issues. *Scalable Comput. Pract. Exp.* **2019**, *20*, 377–398. [\[CrossRef\]](#)

152. Kumar, K.A.; Quamar, A.; Deshpande, A.; Khuller, S. Sword: Workload-aware data placement and replica selection for cloud data management systems. *Vldb J.* **2014**, *23*, 845–870. [[CrossRef](#)]
153. Bonvin, N.; Papaioannou, T.G.; Aberer, K. A self-organized, fault-tolerant and scalable replication scheme for cloud storage. In Proceedings of the 1st ACM symposium on Cloud computing, Indianapolis, IN, USA, 10–11 June 2010; pp. 205–216.
154. Ascó, A.; Leeds, T. Adaptive Strength Geo-Replication Strategy. In Proceedings of the PaPoC '15 Proceedings of the First Workshop on Principles and Practice of Consistency for Distributed Data, Bordeaux, France, 21 April 2015; Volume 155, p. 161.
155. Mansouri, Y.; Babar, M.A. The Impact of Distance on Performance and Scalability of Distributed Database Systems in Hybrid Clouds. *arXiv* **2020**, arXiv:2007.15826.
156. Aslanpour, M.S.; Toosi, A.N.; Taheri, J.; Gaire, R. AutoScaleSim: A simulation toolkit for auto-scaling Web applications in clouds. *Simul. Model. Pract. Theory* **2021**, *108*, 102245. [[CrossRef](#)]
157. Sousa, F.R.; Machado, J.C. Towards Elastic Multi-Tenant Database Replication with Quality of Service. In Proceedings of the 2012 IEEE Fifth International Conference on Utility and Cloud Computing, Chicago, IL, USA, 5–8 November 2012; pp. 168–175.
158. Sharma, U.; Shenoy, P.; Sahu, S.; Shaikh, A. A Cost-Aware Elasticity Provisioning System for the Cloud. In Proceedings of the 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, USA, 20–24 June 2011; pp. 559–570.
159. Maghsoudloo, M.; Khoshavi, N. Elastic HDFS: Interconnected distributed architecture for availability–scalability enhancement of large-scale cloud storages. *J. Supercomput.* **2019**, *76*, 174–203. [[CrossRef](#)]
160. Stauffer, J.M.; Megahed, A.; Sriskandarajah, C. Elasticity management for capacity planning in software as a service cloud computing. *IJSE Trans.* **2021**, *53*, 407–424. [[CrossRef](#)]
161. Mahmood, T.; Narayanan, S.P.; Rao, S.; Vijaykumar, T.N.; Thottethodi, M. Karma: Cost-Effective Geo-Replicated Cloud Storage with Dynamic Enforcement of Causal Consistency. *IEEE Trans. Cloud Comput.* **2021**, *9*, 197–211. [[CrossRef](#)]
162. Vignesh, R.; Deepa, D.; Anitha, P.; Divya, S.; Roobini, S. Dynamic Enforcement of Causal Consistency for a Geo-replicated Cloud Storage System. *Int. J. Electr. Eng. Technol.* **2020**, *11*.
163. Seguela, M.; Mokadem, R.; Pierson, J.-M. Comparing energy-aware vs. In cost-aware data replication strategy. In Proceedings of the 2019 Tenth International Green and Sustainable Computing Conference (IGSC), Alexandria, VA, USA, 21–24 October 2019; pp. 1–8.
164. Khalajzadeh, H.; Yuan, D.; Zhou, B.B.; Grundy, J.; Yang, Y. Cost effective dynamic data placement for efficient access of social networks. *J. Parallel Distrib. Comput.* **2020**, *141*, 82–98. [[CrossRef](#)]
165. Liu, J.; Shen, H.; Chi, H.; Narman, H.S.; Yang, Y.; Cheng, L.; Chung, W. A Low-Cost Multi-Failure Resilient Replication Scheme for High-Data Availability in Cloud Storage. *IEEE ACM Trans. Netw.* **2020**, 1–16. [[CrossRef](#)]